Recognition of Food Type Using Deep Learning and Calorie Estimation

A Thesis submitted to the

UPES

For the Award of

Doctor of Philosophy

in

Computer Science & Engineering

By

Ritu Agarwal

November 2024

Supervisor
Dr. Neelu Jyothi Ahuja

Co-Supervisor
Dr Tanupriya Choudhury

External Supervisor
Dr Tanmay Sarkar

Department of Computer Science & Engineering
School of Science (SOCS)
UPES
Dehradun- 248007: Uttarakhand

Recognition of Food Type Using Deep Learning and Calorie Estimation

A Thesis submitted to the

UPES

For the Award of

Doctor of Philosophy

in

Computer Science & Engineering

By

Ritu Agarwal

(SAP ID:500089264)

November 2024

Supervisor
Dr. Neelu Jyothi Ahuja
Professor & Associate Dean (Academics)
School of Computer Science
UPES

Co-Supervisor
Dr Tanupriya Choudhury
Professor, AI Cluster
School of Computer Science
UPES

External Supervisor
Dr Tanmay Sarkar
Department of Food Processing Technology
Malda Polytechnic, West Bengal State
Council of Technical Education
Government of West Bengal, Malda,India

Department of Computer Science & Engineering
School of Science (SOCS)
UPES
Dehradun- 248007: Uttarakhand

DECLARATION

I declare that the thesis entitled **"Recognition of Food Type Using Deep Learning and Calorie Estimation"** has been prepared by me under the guidance Dr Neelu Jyothi Ahuja, Professor & Associate Dean (Academics) of School of Computer Science, UPES, Dr Tanupriya Choudhury, Professor, AI Cluster, SOCS, UPES and Dr Tanmay Sarkar, Department of Food Processing Technology, Malda Polytechnic, West Bengal State, Council of Technical Education, Government of West Bengal, Malda, India. No part of this thesis has formed the basis for the award of any degree or fellowship previously.

RITU AGARWAL
School of Computer Science
UPES
Bidholi via Prem Nagar, Dehradun, UK, INDIA

Date:

CERTIFICATE

I certify that Ritu Agarwal has prepared her thesis entitled **"Recognition of Food Type Using Deep Learning and Calorie Estimation",** for the award of PhD degree of the UPES, under my guidance. She has carried out the work at the Department of Computer Science & Engineering, University of Petroleum & Energy Studies (UPES).

Neelu

Supervisor

Dr. Neelu Jyothi Ahuja
Professor & Associate Dean (Academics)
School of Computer Science
UPES

Tanupriya Choudhury

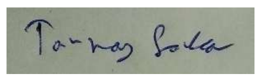
Co-Supervisor

Dr Tanupriya Choudhury
Professor, AI Cluster
School of Computer Science
UPES

Date:

CERTIFICATE

I certify that Ritu Agarwal has prepared her thesis entitled **“Recognition of Food Type Using Deep Learning and Calorie Estimation”,** for the award of PhD degree of the UPES, under my guidance. She has carried out the work at the Department of Computer Science & Engineering, University of Petroleum & Energy Studies (UPES).

External Supervisor 

Dr Tanmay Sarkar
Department of Food Processing Technology
Malda Polytechnic, West Bengal State
Council of Technical Education
Government of West Bengal, Malda,India

Date:

ABSTRACT

Food is crucial for the human body. India is known for its rich heritage, diverse culture, and various Indian dishes celebrated worldwide. Monitoring calorie intake is essential for maintaining good health today, where excessive calories can lead to diseases. Due to the association between bad eating habits and a host of health problems, people are more concerned about their meals. Calorie consumption should be considered for a healthy life in a balanced diet. Measuring the amount of food in each item is a highly challenging assignment for an individual. Nowadays, following a standard healthy diet is essential to prevent obesity. Individuals need a system to guide them on food portions, calorie intake, and consumption directions for a balanced diet. The identification of Indian food cuisine utilizing machine vision-based techniques remains limited. The available Indian data set is insufficient for studying deep learning (DL) classification systems. The research uses a hybrid deep learning algorithm and a unique dataset to identify and accurately estimate the calories in food items.

Previous attempts to find an image dataset suitable for the research needs were unsuccessful, leading to the design of a custom dataset. As the Indian dataset is lacking, the IndianFoodNet dataset was created to enhance machine vision-based techniques for identifying everyday Indian food products. Initially, I compiled a list of commonly consumed food items in India from different regions. After compilation of a list of classes, collected multiple dish images and label each product's location on each image using borders. Then processed more than 5000 images and were tagged with more than 15,000 food items from 30 different cuisine classes, creating the IndianFoodNet dataset. Various YOLO algorithms determine food items' caloric content and train a model for their recognition. A comparative study was provided for numerous state-of-the-art object detection models- “YOLO5, YOLO7, and YOLO8”. The objective is to develop a DL-based detection system that can identify Indian cuisine.

Additionally, the model's performance has been examined and assessed using a qualitative analysis of the predictions generated on the dataset's separated images for testing. (As seen in the summary of training of YOLO at five epochs, YOLO8's precision is greater by 0.775 times in comparison with YOLO7's and YOLO5's

precision). When compared to YOLO5 (recall value = 0.671) and YOLO8 (recall value = 0.719), YOLO7 has the lowest recall value. The results of this study may contribute to the creation of applications that will help Indian food providers and consumers, ushering in a new era of food technology. Additionally, the results may boost the efficacy of computer vision-based methods for recognizing different Indian cuisines.

After inspection and evaluation of the dataset on various YOLO models, the work's final contribution offers a hybrid framework for evaluating food products' calorie content using DL methods. In the framework, start by segmenting the images using the mask R Convolutional Neural Network (CNN). Following segmentation, features are extracted from the divided images, and the YOLOV5 architecture is used to classify the food. When all the items of food have been identified, their approximate dimensions are determined, and each meal product's quantity is computed using these measurements by which volume of food is evaluated. The proposed model effectiveness is measured with various deep learning algorithms, like "CNN, YOLO, and Mask (R-CNN)", based on several metrics, including recall precision, accuracy, sensitivity, MCC, NPV, specificity, FPR, FDR, and FNR. Evaluation measures show that, compared to current approaches, the suggested method has much lower error levels. Comparing the suggested strategy to other models, it shows an overall high accuracy of 97.12% for a single image (taken randomly). When we take 10% of testing images in a loop randomly, then 94% accuracy is achieved, which is high overall, providing a robust and reliable solution for identifying and estimating the content of calorie in Indian food items.

ACKNOWLEDGEMENT

I bow my head humbly to pay heartfelt regards to Almighty God for giving me the strengths and blessing in completing this thesis.

There are quite a few people that have helped me in one way or another in completion of this work. It is with great pleasure; I would like to thank all of you from very deep inside.

Foremost, I would like to express my sincere gratitude to my thesis advisor Dr. Neelu Jyothi Ahuja, Dr Tanupriya Choudhury and Dr Tanmay Sarkar for picking me up as a student at the critical stage of my career and for the continuous support in my Ph.D study and research, for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me throughout the research and in writing of this thesis.It is absolutely difficult to succeed in the process of finding and developing an idea without the help of a specialist in the domain. I found in my advisors not only the source of wonderful ideas to develop, but also the source of support that a Ph.D student needs. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to **thank the Chancellor – Dr. Sunil Rai, the President-Dr Ram Sharma, the Dean of School of Computer Science, the Dean-R&D, the Associate Dean (R&D) of School of Computer Science,** at UPES for their encouragement, suggestions and valuable support through my research journey.

I am grateful to UPES for allowing me to pursue research and providing the facilities at the School of Computer Science. I thank the faculty members and Doctoral Research Committee members, Dr. H. G. Sastry, Dr. Lalit Kane, Dr. Alok Agarwal, Dr. Kaushik Ghosh, and Dr. Varun Sapra, for their support and guidance. I express my gratitude to my family for all the love, support, encouragement, and prayers they have sent me along this journey.

I would like to express my special thanks to Dr. Rakhi Ruhal, Program Manager-Ph.D, and Mr. Yogesh Kumar Sharma, Exam Office – Executive, at UPES, for their assistance during my research work.

I am eternally indebted to my loving parents and in-laws for all the sacrifices they have made on my behalf. I would like to express my sincere gratitude to my elder sister Ms. Neetu for her unconditional support throughout my PhD journey. I would like to express sincere gratitude to my beloved husband Avesh Tyagi who believed in me and provided encouragement during challenging times. Your unconditional support in the moments when there was no one to answer my queries has helped me immensely.

To my children Aniya Tyagi and Rishit Tyagi, who are the inspiration for me to complete this journey and also for the sacrifices made along the way. To my friends Dr Shilpa Srivastava and Mr. Vibhor Agarwal for their support and constant encouragement.

RITU AGARWAL

TABLE OF CONTENT

Title	*Page No.*

Title	*Page No.*

LIST OF FIGURES

Figure No.		*Page No.*

LIST OF TABLES

Table No.		*Page No.*

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

In the introduction of this chapter, problem statement, need and motivation, and research questions have been described. Subsequently, the objective of research, contribution and organisation of thesis have been presented.

In the modern world, individuals must understand what they eat and how it affects their bodies. Consumers choose snacks or sodas that're richer in sugarEating an imbalanced diet and consuming foods high in calories but low in nutrients are the main causes of obesity. *Excess* is an illness that refers to an excess of body fat that is detrimental to the physical body [1]. There is a rise in obesity because individuals are becoming less active, don't know enough about diet foods, and have disordered eating patterns. When someone consumes more food than their body uses daily, they are on the path toward obesity. Obesity is connected to a number of conditions, including breathing difficulties, diabetes, hypertension, and heart problems. Because of their excess body weight, obese individuals suffer from damage to the ligaments in their knees and other joints. Additionally, they have trouble in breathing while walking or climbing stairs, and their heart strains excessively when pumping blood. Overweight and obesity lead to several serious and persistent illnesses. In 2013, the American Medical Association publicly acknowledged obesity as a medical problem that requires attention and has negative consequences on a patient's health [2]. In order to prevent obesity in the human body, a balanced diet and exercise is necessity for regular consumption of healthy foods. Obese people can only lose weight in a healthier way and improve their health if their daily food intake is properly measured. Eating a nutritious, balanced diet and regular exercise will help lower the obesity rate and promote a normal body mass index (BMI) and a healthy life. Losing weight can be helped by eating foods in balance that are low in calories and high in nutrient fibre. It is highly challenging for an individual to know the amount of nutrients in each food and to measure the volume of food. As a result, everyone needs a system or helper to direct them and provide information on the quantity and calories of meals. Having a nutritionist with everyone all the time to advise on food is quite challenging. They require an automated system that can help anybody at any time with a simple meal image and provide comprehensive details about its nutrients and content. The accuracy of calorie computations and food item identification appears to have improved.

Food recognition is becoming increasingly vital in today's society, where there's a heightened emphasis on health, nutrition, and the integration of technology into our lives. Here are some important reasons why food recognition matters and why it's crucial to tackle this issue:

1. Health and Nutrition Awareness

The past few years have seen a surge in global awareness about health and nutrition, fueled by rising concerns over obesity, diabetes, cardiovascular diseases, and other diet-related health problems. Effective food recognition systems can empower individuals to more easily track and manage their dietary intake, leading to more informed choices about what they eat.

- Problem significance:Many people find it challenging to accurately gauge portion sizes and calorie counts of the food they consume, which is key for managing their diets effectively. A dependable food recognition system that includes calorie estimation can automate the process of tracking daily nutritional intake, encouraging healthier eating habits and better control over health conditions.
- Real-world application:These systems are particularly advantageous for individuals who are dieting, managing chronic illnesses (like diabetes), or adhering to specific nutritional plans (such as athletes or bodybuilders). By streamlining food tracking, users can maintain more precise food logs, equipping them to make better health decisions.

2. Obesity and Diet-Related Health Conditions

The increasing rates of obesity across the globe are contributing to a rise in serious health issues like diabetes, hypertension, and heart disease. One key obstacle in tackling these conditions is effectively tracking dietary habits. Food recognition systems can equip users with straightforward, real-time tools that help them keep an eye on what they eat.

- Public health impact: To effectively address obesity and related diseases, there's a need for enhanced monitoring tools to manage food consumption. Traditional methods of logging food intake often take a lot of time and can be quite inaccurate. A food recognition system simplifies this task, enabling individuals to adopt healthier practices and lower their risk of diet-related illnesses.
- Significance for healthcare providers: Incorporating automated food recognition into healthcare and wellness programs can aid patients in managing dietary issues. Nutritionists, dietitians, and doctors can leverage these systems to offer informed dietary guidance and tailored advice.

3. Technological Integration into Everyday Life

As smartphones and wearable devices become ubiquitous, many individuals are turning to technology for monitoring and managing their health and fitness. Food recognition systems align perfectly with this trend towards the "quantified self," where people utilize data-driven insights to enhance their overall wellness.

- Real-world significance: A food recognition app that connects with smartphones or wearables allows users to snap pictures of their meals and receive immediate insights about their nutritional content. This convenient interaction with everyday technology makes food recognition an essential resource for anyone aiming to maintain a healthy lifestyle.
- Convenience and accessibility: Given the widespread use of mobile devices, food recognition systems have the potential to serve millions, offering a highly accessible solution for tracking diet. This convenience removes the traditional hurdles that often hinder food logging, such as limited time and lack of nutritional expertise.

4. Smart Restaurant Menus and Food Delivery Services

Food recognition technology has significant potential in the restaurant and food delivery sectors. As the demand for food delivery continues to rise, automated recognition systems can greatly improve the user experience by offering accurate nutritional information for the meals ordered.

- Restaurant Applications: Restaurants can implement smart menus that leverage food recognition technology, allowing customers to easily access nutritional details by simply scanning their dish. This not only enriches the dining experience but also empowers patrons to make more informed choices while eating out.
- Food Delivery Services: While many food delivery platforms provide nutritional information, it often lacks accuracy or is incomplete. By incorporating food recognition, these platforms can ensure customers receive precise nutritional breakdowns of their meals, making it easier for them to align their food options with their health goals.

5. Cultural and Global Significance

Food plays a crucial role in cultural identity, and food recognition systems can enhance our understanding of global cuisine. Given the diversity of food across various cultures, the ability to recognize and classify different food types can foster better cultural exchange and appreciation. Furthermore, these systems that include regional dishes can expand the availability of nutritional data for varied populations.

- Global Health Initiatives:Food recognition technology can support global health programs aimed at improving nutrition in diverse regions. By identifying local and regional foods, these systems can help monitor eating patterns and promote nutritional education in areas with lower food literacy.
- Food Sustainability: In terms of sustainability, food recognition can aid in minimizing food waste by accurately identifying portion sizes and guiding consumers toward more efficient food choices.

6. Integration with Emerging Technologies

The rapid advancements in AI, augmented reality (AR), and the Internet of Things (IoT) present exciting opportunities to integrate food recognition into emerging technologies. For instance, smart kitchens powered by IoT devices could utilize food recognition for recipe suggestions, automated grocery lists, and nutritional tracking.

- IoT Applications: As smart home devices become increasingly prevalent, food recognition may become a core feature in the next generation of kitchen appliances. For example, refrigerators equipped with cameras could utilize food recognition to track inventory, suggest recipes, and notify users about expiring products.
- AR Integration: AR applications can elevate the dining experience by overlaying nutritional information, portion size estimates, and even cooking instructions onto real-world food items through smartphone cameras or AR glasses.

7. Support for Individuals with Visual Impairments

Food recognition technology holds great promise for those who are visually impaired. By employing image recognition to distinguish various food items, these systems can assist users in navigating restaurant menus or grocery stores, providing spoken information about the food and its nutritional details.

- Accessibility benefits: This innovation could greatly enhance the independence of visually impaired individuals, enabling them to engage more confidently in food-related tasks, from meal preparation to making informed choices about healthy eating.

8. Insights for Research and Data Gathering

Effective food recognition systems can yield valuable insights for extensive nutritional studies. By gathering comprehensive data on real-world food consumption habits, researchers can gain a deeper understanding of dietary patterns, nutritional shortfalls, and their implications for public health.

- Importance for nutritional studies: Through analyzing the data obtained from food recognition systems, researchers can reveal trends in eating behaviors across populations, evaluate the success of public health initiatives, and pinpoint developments relevant to diet-related health issues.

1.2 FOOD COMPUTATION

Mostly Computer science methods are for food computation in food-related studies. The five fundamental tasks of food computation are awareness, identification, retrieval, projection, and monitoring suggestions. It has various applications in numerous industries, including agriculture, medicine, tradition, and fitness. Data about food is claimed to be gathered from several data sources during the first phase. Applying data mining, computer vision, machine learning, and other techniques for analyzing food data using food databases. Identification is a job mostly used to predict food goods from food images, including the components or categories [3]. Retrieval-based approaches for food recognition are also feasible when the food image categories are large. Online estimates can prove useful for making suggestions as well.

1.2.1 Acquisition of Food-data and Analysis

We can get the person's meal log data for analysis to assess their eating patterns for their dietary review. Therefore, acquiring and compiling food images is the initial stage in the food calculation process. In order to calculate calories in food, consumers may use the nutrition formula offered by food organizations such as the United States Department of Agriculture (USDA4) and BLS5. Researchers get data of food from official organizations and specialists for conduction of food analysis. [4] For food image recognition, food potting. com's first developed large-scale Food-101 dataset. Internet-based platforms like Twitter, Facebook, Foursquare, etc., and sites that publish dishes, provide many foods information [5]. Gathering information about meals via IoT devices is also a widespread practice.

Various forms of data are available for studying food identification and calorie estimation.

Food Images: DL techniques may be applied to understand food through easily accessible images of meals.

Cooking Videos: Various forms of data are available to study food identification and calculate caloric content:

Food Characteristics: Knowledge of ingredients, cooking methods, flavours, colours, and tastes improves the ability to identify foods and retrieve recipes.

Food Logs: Applications such as Food Log facilitate food intake monitoring by offering nutritional and calorie information to support a balanced diet.

Health Data: This data provides dietary details and calories, which is centred on health and crucial to managing weight and avoiding health problems.

Additional Food Data: This group includes data from government cookbooks, surveys, databases of order levels, and food product codes.

Figure 1.1: Sample food item images from datasets

In Figure 1.1, the first two rows contain images from the Yummly-66 K (Indian class) food dataset, the next two rows are from real-time foods (images captured at hotel and home), and the last two rows contain images from web sources.

1.2.2 Application

Health

Our well-being depends on what kind of food we consume or how much we consume. For example, if we overeat, we could be at risk of various diseases, like diabetes, obesity, and heart disease [6]. Food-relevant studies would also benefit from several health-specific applications. Figure 1.2 shows some of the images where food images are used related to health applications.

Figure 1.2: Food images in health applications

Following food-oriented health applications are also added: (1) understanding the food for health, (2) diet control, (3) food suggestion, and (4) social media food-health research.

Food Understanding for Health

Lots of studies have researched how we view food, both before and after eating, due to the overweight and obesity epidemic. The emphasis [7] was on how multi-modal signs (perception and aroma) influenced food recognition and food choice guidance.

Health Diet Control

The Food Diary or Dietary Assessment [8][9] offers valuable insights into the prevention of most diseases. Conventional approaches rely mainly on opinions or self-reporting. However, with ―underestimation and miscalculation of food intake, these methods have several issues. Many processes retreat to perception-based strategies for dietary control with the growth of computer vision strategies. With this knowledge, the first digital imagery tool was developed for food intake analysis [10], which assessed food intake in restaurant environments. A computer program instantly evaluates the grams and kilo calories of a particular food portion size (based on the USDA database).

Health-aware Food Suggestion

They commonly discussed Health concerns and changing eating habits when motivating research on food advisory systems. For example, in the recommendation algorithm, [11] took calorie counts into account on the basis of their suggested "calorie balance function" which gives comparison between calories that the consumer requires and those from a recipe.

Social Media Food-Health Research

We are in a social forum age, as food is essential to our lives. Users upload images of food, discover related recipes, and chat about them. The majority of data regarding our eating patterns, gastronomy, and conduct is captured on social media.Current findings suggest that we use social platforms for public health tracking to collect aggregate data on people's health. A large study [12] presented 27 figures for health statistics related to well-being, such as healthcare coverage and obesity.[13]. Gather food in many tweets, which were used to forecast health parameters, like overweight and diabetes-type health problems.

Obesity a disease

In India, as well as other countries, obesity is a severe health issue. It is a phrase used to characterize someone who is significantly overweight. Obesity may lead to chronic illnesses, including diabetes, heart disease, colon cancer, and hypertension, all of which may be detrimental to an individual's welfare. Obesity is mainly caused by those who ingest more significant amounts of fat instead of exercising through physical activity (PA) [14]. Adults are considered obese if their BMI is more than thirty. A child's BMI is considered obese if it is higher than the 95th percentile for their gender and age, based on clinical criteria [15]. Several factors can cause weight. One of these explanations is that burning many calories. According to statistics, over the last 25 years, the rate of obesity has climbed from 2.3 to 3.3 times. Children who are obese have a higher chance of acquiring a number of chronic illnesses, including high blood pressure, insulin resistance, and type 2 diabetes [16, 17]. Approximately thirty percent of kids are overweight or obese, according to current statistics, and research indicates that kids are getting fatter earlier in life and continuing to remain fat over time [18]. Figure 1.3 demonstrate how fat is on the rise in society and how it impacts people of all ages.

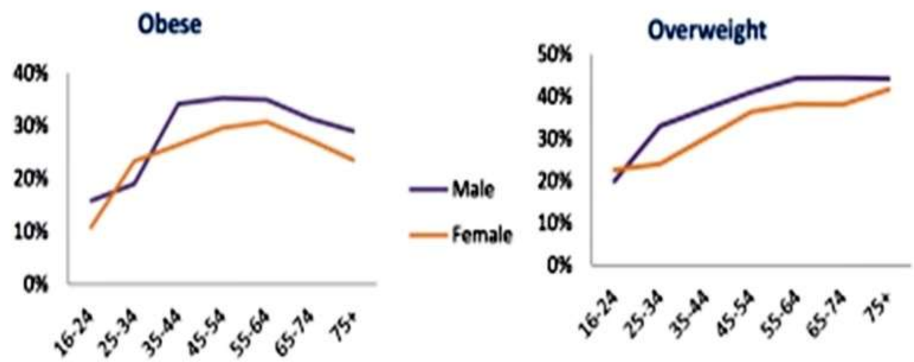


Figure 1.3: Trends in adult obesity and overweight across various age groups

Diet's Effect on Health and Society

These days, obesity is on the rise in both adult and paediatric populations, and it has risen quickly to the point that it is one of the world's leading preventable causes of mortality [18, 19]. The World Health Organization (WHO) states that a high body mass index (BMI) increases the risk of developing several chronic conditions, such as musculoskeletal issues, malignancies of the breast, ovary, liver, and kidney, and cardiovascular diseases like heart disease and stroke. Additionally, obesity hurts society at large and strains medical facilities across the world. Research from 2016 estimates that the National Health Service (NHS) bears direct expenses of £6.1 billion a year due to obesity; by 2030, costs will rise by an additional £2 billion. The Association of Indirect expenses might be poor health, lost income, or health issues brought on by obesity, which might involve lower productivity and cost companies' money if employees miss work because of high BMI [20]. One crucial strategy (together with physical exercise) to combat obesity and encourage healthy living is nutritional intake management.

1.2.3 Why Dietary Assessment?

The WHO reports that throughout the past 30 years, the rate of obesity has more than doubled. The most compelling evidence connecting nutrition to an increased risk of cancer is linked to overweight and obesity, as well as to the consumption of fermented foods, aflatoxins, and excessive alcohol. Increased calorie consumption is the leading cause of the rise in obesity [21]. Regularly consuming 400 gram or more of fruit and vegetables decreases cancer risk. It is the result that promoting better nutrition can prevent disease and is a worldwide health priority [22]. Evaluating dietary intake is essential to proving the connection between food and illness. According to studies, persons who track their food consumption or keep a food diary are more likely to lose weight than those who don't [23].

Food Record

Observing school breakfast and lunch [24] is the most popular way to assess the correctness of the 24HR with kids. A meal record is the preferable dietary evaluation approach for clinical investigations, even though 24-hour recall is effective in population-based studies. Due to the difficulty of recording meals, this technique has several significant drawbacks, including the potential for underreporting [25]. Nevertheless, this approach needs people who are highly motivated and literate. A research gap in nutrition and health is the precise assessment of food consumption.

1.2.4 Digital Interventions

Teenagers are expected to own smartphones at a rate of 78.9%, indicating that technology has become ubiquitous among them. This creates a chance to develop programs that encourage physical activity and a healthy diet using smartphone technologies.

mHealth

We may target these people with mHealth applications that can use the built-in devices in mobile devices to support nutritional control because of this technology adoption rate. Many studies have examined the use of smartphones to manage aspects of lifestyle, exercise, and nutrition to promote wellness. For instance, to support weight control, a smartphone application was created in [26] incorporating strategies for altering behavior, such as personal tracking, establishing objectives, and an evaluation phase. The application uses the user's private data, such as size and weight. To calculate daily energy objectives. A daily food consumption log is available to the user. Then, it decreases the whole energy allotment from this intake. An additional mobile application has been created to track the user's eating patterns and exercise routine [27]. The application may guide people on how to lead a healthy lifestyle. In addition, Hutchesson et al.'s recent study investigated the utility of smartphone technology compared to website- and paper-based techniques for self-monitoring nutritional consumption. The findings indicate that, in comparison to conventional paper-based techniques, techniques based on cell phones and computers are famous for regulating food consumption [28].

1.2.5 Dietary Management Through Food Logging

They generally acknowledge that food logging, also known as food journaling, is crucial to encouraging weight reduction or maintaining current weight. A self-management technique for tracking the consumption of nutrients is food logging, which consists of documenting dietary consumption. A great deal of studies is conducting on creating food-tracking methods in order to guarantee user convenience and enhance usefulness. A database application programming interface (API) that lets users look up calorie content is a feature of contemporary meal recording programs. However, this could not be true and might not accurately reflect the food the user has consumed. If a person cannot recall from memory the specific foods they have consumed over a day, this would cause overestimates or underestimates in food logs, impacting overall dietary control. Another significant difficulty is accurately determining the number of calories in a meal piece; consumers reported this in [29].

Food Images for Dietary Management

Images have been used more often in recent years to track food intake for nutritional control. People may examine portion proportions in more detail and learn more about food. A number of image food monitoring treatments ask users to snap pictures of their meals before and after to enhance objectivity and precision concerning the quantity of food consumed. Due to the widespread use of smartphones, individuals may record their food consumption using their high-pixel cameras and do in-depth contextual analysis of their food intake. The basis of these findings is on studies done [30], which show that meal monitoring methods utilizing electronics retain consumers longer than techniques relying on print memos. In [30], the obstacles of food logging are discussed.

In addition, people are using food images on social networking platforms like Facebook, Instagram, and Food Spotting. This practice allows people to share their meals with others and may encourage them to pay more attention to the quality and quantity of their meals. Obtaining a visual breakdown of dietary consumption is another benefit of uploading food images to social media. The Instagram tile interface is displayed in Figure 1.4. Research has indicated that calculating the calories in individual meal items rather than the meal as a whole can improve accuracy [31]. Social cognitive theory was applied in [32] to persuade users to make healthier decisions.

It's evident that there are numerous benefits to using images for diet management; yet, users must still identify the type of meal and look up nutritional information in order to finish the picture food tracking process. Current research has focused on automated methods to identify food intake from images to encourage application adherence and convenience for weight control.

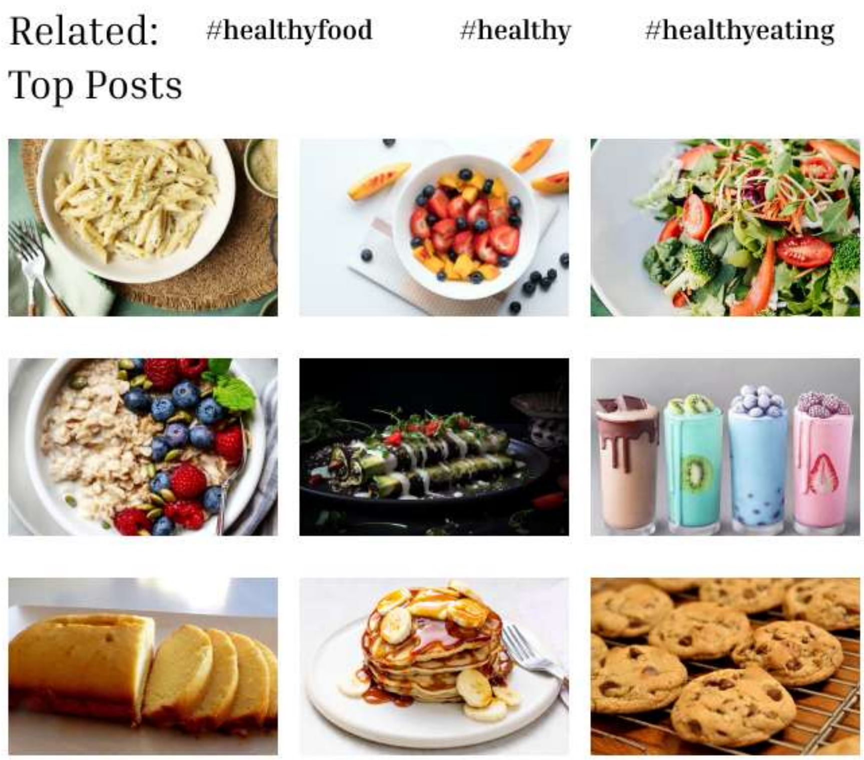


Figure 1.4: Images uploaded to Instagram depicting food meals.

Automated Image Logging of Food

Automated food image logging presents a promising option for food recording. This method involves searching food item names in an online database to obtain nutritional information, and it is a critical component of automated systems for food image recording. Computer vision systems can be applied in various contexts to provide continuous meal image recording by predicting the meal products included in the image and using image processing methods. The potential impact of this automated method on dietary control is significant, as it can streamline the process and improve adherence to weight control applications.

The traditional method of automatic meal tracking involves several stages. First, visual characteristics, such as colour, texture, and form, are extracted from food image collections. Next, the food groupings' most representative traits are chosen. After that, tagged features are taken out of a training image dataset and used to train classifiers for machine learning under supervision, like Support Vector Machines (SVM). Test photos can be classified using the classification model once training is finished. This process requires the same feature or combination of features to be extracted for a trained classifier to classify test images. The image classification pipeline used for automatic food tracking for nutritional management is shown in Figure 1.5. A significant amount of research on applying computer vision techniques for categorizing food images for dietary control has been published in recent years [33].

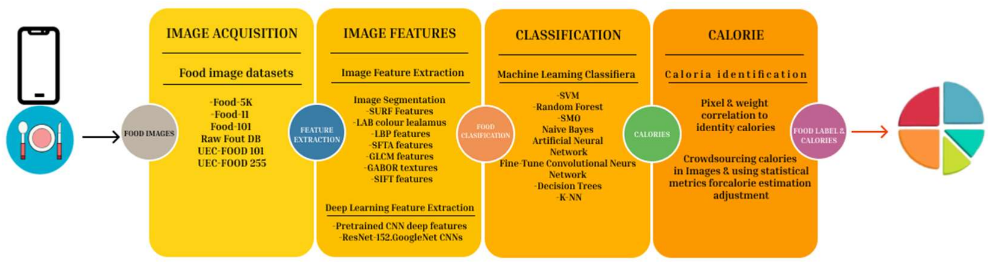


Figure 1.5: Methods that have been employed in earlier studies for automatically logging food images

1.2.6 Effects of Nutrition Transition in India

Changes in Pattern of Food expenditure and future Calorie Intake

It's evident that there are numerous benefits to using images for dietary control. However, users must actively choose the meal type and look up nutritional information in order to participate in the visual food monitoring process. Recent research has been focused on developing automated methods for determining nutritional intake from images, with the aim of promoting adherence to convenience and weight control applications.

Patterns of Food Consumption and Dietary Variability

Changes in food consumption patterns can have positive and negative outcomes in terms of diet quality. They balance the increase or decrease in the consumption of cereals and pulses or calorie intake in the overall diet. The best way to examine the quality of the diet is to use dietary diversity indicators. Using household dietary diversity, it measures diet quality and understands how the nutritional needs of households are being met [34]. A higher-quality diet decreases the risk of obesity and other chronic disorders in children [35]. Suggested that households with low dietary diversity are likely to consume less per person and lower caloric availability.

Dietary Diversity at State and Individual Level and nutritional status of children

Moms who maintain a healthy weight are more inclined to limit their kids' food consumption [36] mentioned that when a kid's weight grows, parents try harder to regulate and limit what they feed them, especially if the child is a girl [37]. It also agreed that dietary diversity in young children is associated with improved nutritional status. Studies have conclusively established that breastfeeding significantly reduces the risk of being overweight in childhood. Breastfeeding benefits infant and child health because it has immunological properties. As a result, enhancing breastfeeding and utilizing other feeding techniques will benefit kids' health, nutrition, and overall well-being both now and in the future [38]. It was shown that following dietary guidelines rich in protein and carbs may lower a child's risk of stunting.

Identifying Food and Calculating Calorie Values

The key to a human body is food. Dietary consumption is a growing concern for individuals since poor diets can result in many ailments. A diet plan must always account for the calories required to ingest to maintain good health and fitness. If a person's BMI is more significant than thirty, they are overweight [39]. A disparity between the number of grams of meals ingested and the health they contain is the leading cause of fat. The majority of people on earth live in societies where obesity and overweight are the leading causes of death over all other illnesses. People must watch their calorie intake to lose weight or get in shape.

Food Recognition

In the computer vision discipline of "food recognition," algorithms are created to identify and sort food items in images or movies automatically. The primary goal is to create systems that can accurately recognize different types of foods from visual data. Similarly, food images and films with descriptions are necessary for computer program training before we can develop food recognition algorithms. The National Nutrient Database (NND) [40] and the Food and Nutrient Database for Dietary Studies (FNDDS) [41] are publications of the USDA. However, because of data quality and gathering techniques, these databases are best suited for human interpretation rather than machine analysis. This fact has spurred the creation of a fast-food image, video, and data database. Creating and testing food identification programs for obesity research begins with building a food database.

How Food Recognition Works

DL Models: CNNs are good at extracting hierarchical information from images and are the foundation of many food identification systems.

Large Datasets: To train such models, a lot of images of different foods are needed, frequently annotated with nutritional data or equivalent categories.

Feature extraction: During training, the DL model learns to automatically extract pertinent characteristics from images, which helps it differentiate between various food types.

Classification: Once trained, the model can classify input images into specific food categories, allowing for automatic recognition.

Use of Food Recognition

Food recognition is essential for dietary monitoring. It enables users to track their daily food intake through images effortlessly, which also facilitates nutritional analysis and helps assess dietary habits. Integrated health and fitness apps and food recognition provide users with valuable insights for informed decision-making about their diet, promoting a more conscious approach to nutrition.

Challenges in Food Recognition

Variability in Appearance: Foods can vary widely, even within the same category. This variability poses a challenge for accurate recognition.

Complex Meals: It can be difficult to identify specific foods in a complex dinner that includes several meals in one shot.

Cultural and Regional Variations: Food appearance can vary based on cultural and regional differences, making it challenging to create universally applicable models.

1.2.7 Food Volume Estimation

Precise assessment of meal volume, or serving size, needs to be revised in medical and nutritional research. Technological advances allow food images to be readily taken, saved, and sent to a nutritional evaluation system. Smartphones with excellent image capabilities are now widely available. The calories of the food are then immediately estimated by analysing images associated with the dish. Calculating how much food is in the image is a crucial stage in this procedure—a method for estimating meal volume requiring two stereo images. The food item's 3D structure was reconstructed as a 3D point cloud that splits into many 2D wedges. After calculating the amount inside each wedge, calculate the capacity of the entire meal item.

Calorie estimation

Calorie estimation refers to determining the energy measured in calories in a particular food item or meal. Calories are a unit of measurement for the energy content of food, representing the energy the body can derive from consuming that food. Accurate calorie estimation is essential for various health, nutrition, and fitness reasons.

1.2.8 Technological Approaches for Dietary Evaluation

Technology-assisted dietary evaluation is thought to be less time- and labour-intensive and may increase adolescent and child compliance. This age group may be encouraged to adopt better behaviours, including self-monitoring and using current technology like smartphones and other portable devices, especially while using the text messaging as Short Message Service (SMS). It was demonstrated that sending daily SMS messages to a sample of young Type 1 diabetic patients improved their self-efficacy, self-management, and medication adherence.

Dietitians in practice and researchers would benefit immensely from nutritional assessment software that leverages mobile phones to gather dietary data. The current technology-based approaches are image-based [44-45], video-based [46], web-based [43], mobile and web-based [44], and mobile phone-based [42]. Sun et al. [47] measured the camera's length and the separation between a particular pair of points and a specific portion of the food's surface using a calibration card. The meal volume is then estimated using these criteria. A movie-based algorithm was created to assess food items on a table arbitrarily depending on their sizes [46]. This approach computes the height, breadth, and length using geometric relationships. However, capturing movies can be challenging in particular circumstances due to its high memory usage and increased pressure on the user.

The challenges with Food Recognition

A precise assessment of food and nutrient consumption is crucial for tracking patients' nutritional status. Food products must be accurately classified for a system like this to guarantee that the right data is accessible for implementing intervention programs for the prevention of obesity. Therefore, in order to improve the intervention programs, there is an increasing demand for precise food item identification and classification.

The present mobile phone-based nutritional evaluation system encounters the problem of misclassification in the automatic recognition of food items because of their diversity and noticeable variations in appearance. As a result, the evaluation of dietary intake may not provide accurate information. A technology-assisted dietary assessment (TADA) system based on mobile devices was developed. It is not easy to automatically recognize food items in an image. Given that few food products, like butter and margarine, have striking similarities, it is evident that identifying food based simply on one feature of the images is quite difficult. Nonetheless, the capacity to access the food object at any size and in any orientation is thought to be necessary for efficient food identification since food is malleable and experiences considerable appearance changes. For instance, depending on what the customer chooses, an apple on a dish may seem small, enormous, or angled differently. As a result, the apple in the user's meal could not be identical in size or orientation to the sample apples in the database. Moreover, it is impractical to anticipate TADA system users to consistently take pictures of food at a specific size or angle. A large number of the foods we eat on a regular basis have a noticeable texture. Present food recognition methods presume that the food images will be seen from the same angles and at comparable scales. However, as it is challenging to guarantee that a user of a nutritional assessment system would align the camera's size and orientation to record food photos in a way that is equal to the target meal images in a database, genuine implementations do not match this requirement. Calculating the serving sizes at a meal from a mobile device photo is one of the many challenges with nutritional assessments based on images. Researchers are dealing with a challenge that gets harder and harder as the need for accurate food item classification to assist intervention programs increases.

Existing Food Recognition Approaches

Initially, it takes the compact characteristics from food images of recognized products stored in a database in automatic food categorization systems. The image has numerous attributes, such as colour, texture, edge position, and orientation, that can be used to extract the features utilized in this procedure.

Feature fusion, segmented-aided food categorization, paired statistics, and supervision-based methods are a few examples of prior techniques for autonomous food recognition. Yang et al. [48] developed a technique for identifying food using pairwise statistics of the image's pixel composition. The main disadvantage of this strategy is that many meals need pairwise pixels. Zhu et al. [49] achieved segmentation-aided food categorization using only global information from colour and texture. However, a realistic system cannot gather ground-truth information about the food objects due to the wide variations in food's size and shape. These ways are expanded upon by Bosch et al. [50], using a mix of both regional and global features. Nevertheless, this method is not entirely automated because choosing a feeding location by an untrained observer would add to the user's workload. Kong et al. [51] improve classification accuracy by employing joint distributions of multi-view perspective images, lessening the uncertainty around food appearances.

1.2.9 Deep Learning Network

A Deep Learning network is a network of linked nodes, or neurons, that process information. It is an electronic system based on human thinking. The field of Deep Learning is relatively young. It gains knowledge from data through training and modifies weights and biases to provide precise predictions.

Concept of Deep Learning in the Food Recognition Model

An example involving food identification might further clarify the idea behind the Deep Learning neural network model. It uses several hidden layers, unlike the shallow computer vision method. Every pixel in the image is examined and used as a data source. The network analyses each pixel's properties at different levels—a Deep Learning model with the network attempting to categorize the food item as an apple in Figure 1.6. The network divides this query into minor inquiries to address the many aspects of Apple based on its training. Here, the food item's colour, texture, shape, and size are among the details about the system inquiries, which makes it possible for the system to accurately classify the food object by identifying its colour, shape, texture, and size, among other food properties. It produces trained model files specifically for food images to categorize them further. The food object is classified into two categories using the Deep Learning approach. Following the deep neural network's training on food images, the model file was created, segment, and featured, and it was retrieved and added to the deep network's hidden layers.

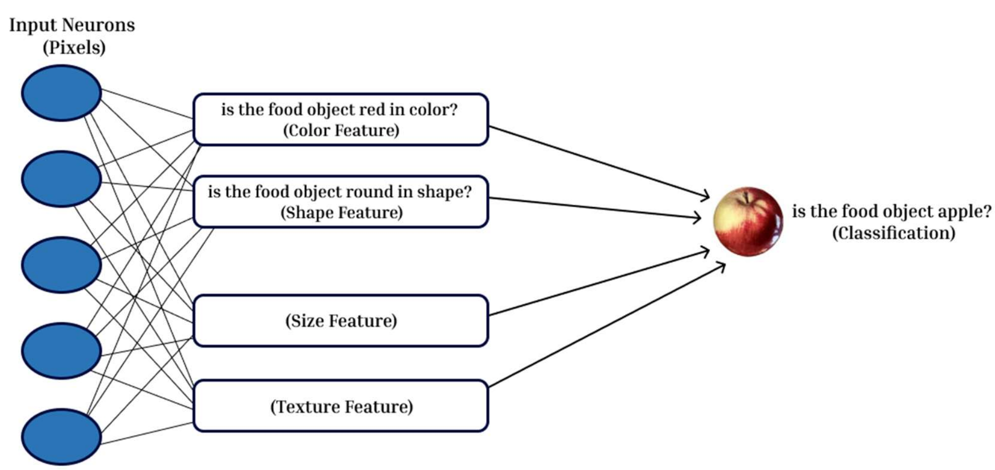


Figure 1.6: Concept of Deep Learning in the food recognition model

Role of Deep Learning in food detection and calorie estimation

Data Gathering

It gathers information (from many sources, including real-time user-taken snapshots, restaurant images, and online food records) using a hybrid Deep Learning technique for food recognition and calorie calculation. Established datasets like Food 101 and Yummly 66K are also used to ensure a diverse and comprehensive collection.

Pre-processing

To ensure the acquired images are suitable for analysis, improve their quality by removing any unwanted noise, adjusting their size, and ensuring a consistent format. This step helps better prepare the data for subsequent analysis.

Food Objects recognition:

Segmentation is applied in the images to separate the food items from the background, which defines the bounds of each food item and assists the system in precisely identifying it.

Food Image Segmentation

Based on a homogeneity criterion, the process aims to segment food images into several relevant food areas (sets of pixels). Different methods work well with different kinds of images. Because there might be a lot of "correct" segmentation for a single image, it can be challenging to evaluate the quality of an algorithm's output qualitatively [52].

Extracting Important Features:

The segmented images of food are the source of crucial characteristics that the hybrid DL model extracts. These characteristics include the food's form, colour, and texture, essential for differentiating between foods and calculating their calorie content.

1.2.10. Calorie Estimation

For the detected meal, the algorithm estimates the calorie value of the identified items using the extracted attributes, incorporating information from various sources to make the predictions more accurate.

Finger-based calorie measurement

Use the user's thumb, where it is on the plate for the finger-based calorie estimation, as seen in Figure 1.7. It is measured once during a single validation process, offering an exact comparison for determining how significant meal portions are in real life. Figure 1.7(c) provides examples of measuring and isolating the thumb and taking food images. The thumb was more adaptable, dependable, and controlled than the calibrating techniques employed in comparable systems. The user found it challenging to take an image with one hand on the phone and the thumb of the other hand close to the plate, making the calibration problem problematic.

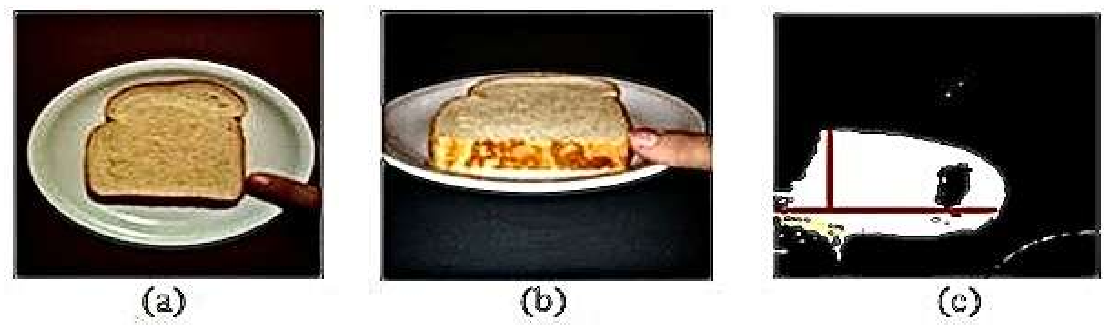


Figure 1.7: (a, b) Test images with food and thumb; (c) Calculation of the thumb dimensions

So, it’s important to create a technique to determine the distance between the thing on the plate and the person taking the image to handle this issue. When the user is going to take an image during the live camera stream, the program determines how far they are from the food item. The training of the Deep Learning model is to correlate input images with particular food groups using a variety of neural networks. The more data it can use to understand and gain knowledge, the more accurate the model will be. The Role of Deep Learning in recognition of food and calorie estimation is shown in Figure 1.8. The user upload image of food, then image is input to deep learning model where image acquisition, preprocessing, object detection and recognition of food take place.

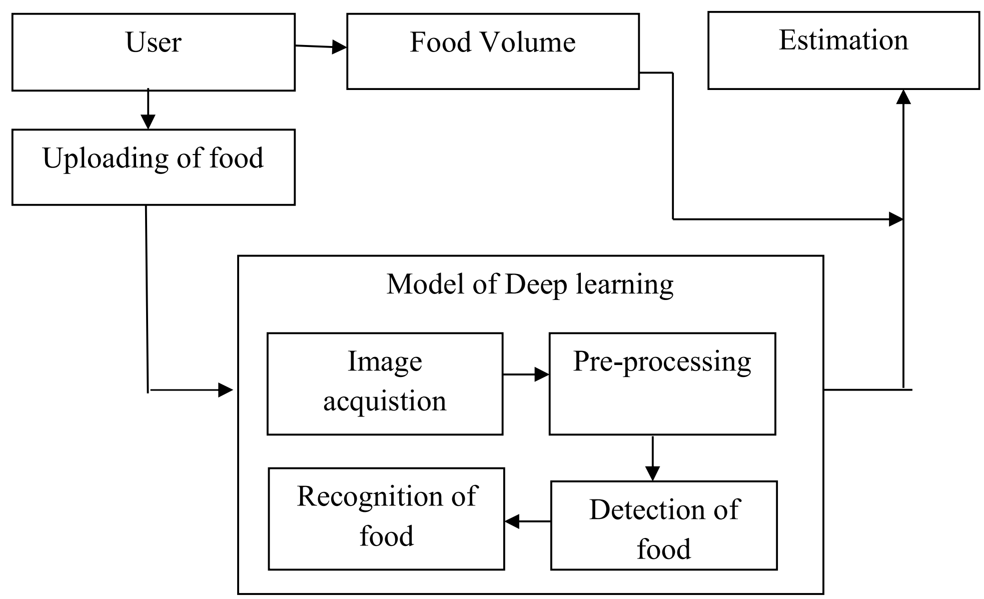


Figure 1.8: Role of Deep Learning in recognition of food and calorie estimation

1.3 PROBLEM STATEMENT

The primary goal of this research is to create an intelligent mobile application that analyses caloric consumption and helps users maintain a well-balanced diet. The application's objective is to figure out the calories in a meal, depending on the location of the mobile device in relation to the food, taking into account differences in how much each food item contributes to the total number of calories. We need technology that can precisely quantify the calories in food in addition to being able to identify them with exceptional accuracy.

Furthermore, current algorithms need a dedicated strategy to address the volume of the intake of meals, particularly in the context of plate-based consumption. The variation in plate size and the arrangement of food items can significantly affect the overall calorie estimation process. Without a systematic approach to account for these factors, users have unreliable calorie counts and inadequate guidance for maintaining a balanced diet. However, existing calorie-calculating algorithms lack methods to account for distance and size, crucial factors for estimating the scale of a food object. The largest of these flaws is that the distance between the user and the food items being analyzed is not taken into account. If this element is not considered, current algorithms are prone to errors that result in inaccurate nutritional evaluations and misleading recommendations.

A more efficient process is to enhance the precision of calorie counting and eliminate the need for users to consult listings manually. Current research aims to address identified issues with present calorie-calculating algorithms, particularly regarding distance and volume of intake meal (plate) strategies, to improve efficiency and processing time.

1.4 RESEARCH MOTIVATION

Over a long time, there has been a notable increase in health issues, with a nearly threefold increase in specific significant problems since 1975. A startling 39% of people over the age of 18 were overweight in 2016, and 13% of them had worries about their obesity (WHO, 2018). These issues are conditions such as cardiovascular problems, hypertension, and menstrual irregularities. As society becomes health-conscious, individuals increasingly focus on managing their diets to address these health challenges. Many people adopt various strategies to lose, gain, or maintain weight.

While these contemporary methods are convenient for initial calorie counting, the results are based on generalized portions that may not accurately reflect the individualized quantities users prefer. This introduces inaccuracies as users' names have their own portions. This issue may be resolved using a formula for figuring out the real quantity of foods that gives accurate and credible calories given desires. The goal is to refine calorie counting by improving result accuracy and eliminating the need for manual table lookups during meal planning, ultimately assisting individuals in creating more dependable diet plans.

1.5 RESEARCH QUESTIONS

Distance Not Considered: Existing algorithms fail to incorporate the distance between the user and the meal item as a determining factor in calorie calculation.

Volume of Intake Meal (Plate) Strategy Not Proposed: The current algorithms lack a specific strategy to account for the volume of the intake meal (plate), which is crucial for accurate calorie estimation.

The problem statement revolves around the need to refine existing calorie-calculating algorithms by addressing these critical issues. The goal is ultimately to achieve a more accurate and user-friendly approach to calorie counting and diet planning.

Current research aims to develop an algorithm to divide up food's portion sizes, determine the appropriate amounts, and then, using the values, produce more precise and trustworthy calorie information. For the same, framing a few of the following research questions as:

RQ1. What are the image segmentation techniques, and which is the best model to use?

RQ2. How do we measure the distance of food objects with the image capture?

RQ3. Which techniques should be best practices in training the model to obtain the precise and reliable result?

RQ4. Which algorithm can detect overlapped objects?

RQ5. In order for the model to retrieve the data during the inference run and display the correct number of calories according to the count, how can we create a sufficient database including the calorie information for each class?

RQ6. How to integrate a web application with a basic user interface using the learned detection model.

RQ7. How can the accuracy and reliability of calorie counting be improved?

RQ8. How can the object detection algorithm's error and detection failure rates be decreased?

RQ9. How can an appropriate approach be found to segment all the classes into smaller portions to achieve accurate calorie counting for a simple plated meal?

RQ10. How to convert 2D image to 3D using image caliberation?

1.6 RESEARCH OBJECTIVES

1. To develop a model for recognizing food images using deep learning algorithms based on the volume of food and estimate the calories of the recognized food images.
2. To improve the reliability with the premise of accuracy and processing time of the existing calorie estimation techniques in food images.
3. Deployment of the proposed model in real time scenario through web-based interface.

1.7 CONTRIBUTION OF RESEARCH

The proposed work offers three main contributions:

Phase 1

- To create a precise detection algorithm for accurately recognizing specific food items using DL techniques.
- Here, the IndianFood-7 Dataset is provided to enhance machine vision's ability to detect common Indian foods. The dataset includes over 800 images annotated with 1700+ labels, covering seven different types of Indian cuisine.
- Various techniques for picture augmentation were applied in order to increase the model's resilience.
- Metrics including precision, F1 score, recall and mAP were used to assess the effectiveness of two object recognition techniques, YOLOv5 and YOLOR, using a test dataset that contained several variations of Indian cuisine items.

Phase 2

- Introducing the IndianFoodNet dataset, which includes over 5500 excellent images and over 5000 annotations covering thirty Indian cuisine items.
- This would enhance the efficacy of computer vision-based methods for recognizing Indian cuisines and streamline the creation of applications that assist Indian food merchants and patrons.

- This work presents a comparative analysis of three cutting-edge object identification models: YOLO5, YOLO7, and YOLO8.
- Also, qualitative studies of the projections built with the dataset's images, which were split to test, examined their efficacy.

Phase 3

A hybrid architecture utilizing DL algorithms is suggested to anticipate the amount of energy in different foods in a vessel.

- This has three primary elements: the division and categorization of food items and the calculation of their volume and caloric value.
- First, mask (R-CNN) is used for segmentation of images.
- The food item is classified after characteristics are gathered from the segmented images using the YOLO V5 framework.
- After identifying meal products, measurements for every meal product are ascertained. Using the predicted dimension, the meal products quantity i.e volume is calculated. Next, the volume of the meal product is used to estimate the energy.
- A variety of metrics, including sensitivity, specificity, accuracy, NPV, recall, F-score, MCC, and FNR, were used to evaluate and analyze the effectiveness of a suggested method on a number of categories, including the CNN network, YOLO, and Mask (R-CNN).

1.8 THESIS ORGANIZATION

The following six chapters comprise the entire thesis.

Chapter 1, includes introduction of research title, problem statement, research motivation, research questions, research objectives and contribution of research.

Chapter 2, details the background and literature review on food recognition using machine learning, food recognition using deep learning, food calorie estimation, food defect detection and conclusion of survey.

Chapter 3, describe the design of proposed methodology, the algorithm and techniques used, and the flow of the model framework.

Chapter 4, implementation of a hybrid architecture to assess the calorie content of different food items within a lunch bowl by utilizing deep neural networks. Three essential parts make up this architecture: volume and calorie computation, segmentation, and classification.

Chapter 5, evaluates and develop a system capable of effectively detecting food and estimating their calorie content. It also covers the system's simulation parameters and the assessment metrics that are used to examine its operation. Moreover, thorough explanations of the experimental outcomes obtained in this study are presented, along with an analysis of the model's performance.

Chapter 6, covers the summary of research contributions, conclusion of thesis, recommendations and future scope.

CHAPTER 2
LITERATURE REVIEW

2.1 OVERVIEW

This part thoroughly discusses various techniques for detecting food and estimating calorie content. It delves into the detailed exploration of DL and machine learning methods employed in these tasks—section 2.2 covers machine learning techniques for recognizing food. Section 2.3 discusses DL approaches to recognizing meal images—calorie estimation methods for food in Section 2.4. As a result, Section 2.5 occupies the detection of food defects.

2.2 FOOD RECOGNITION USING MACHINE LEARNING

Two essentially distinct classification tasks were taken on by Young Im Cho et al. [53]: (i) visibility estimation and (ii) CNN-based food identification. For every activity, they devised two unique data-driven methods that operate in real-time using camera images as inputs, cutting down on calculation time and expenses. Using a recently assembled collection of images captured by closed-circuit television (CCTV) cameras under various weather conditions, such as low- clouds and dense fog, the first approach was designed to predict visibility. On the other hand, the second model used synthetic images to identify dishes. They created a fictitious dataset to improve model performance by gathering a few images from the internet and using data augmentation methods. Both models showed good classification accuracy and low processing time and power requirements.

Various methods developed for the recognition and categorization of images of agricultural and horticultural output were described by Dayanand G. Savakar et al. [54]. They created a BPNN-based classifier that distinguished and categorized various agricultural and horticultural items using colour, texture, and morphological parameters. Both the linear and quadratic discriminant functions were used in the classifying process. Even though each of these parameters offered a varied level of accuracy for various cereals, mangoes, grains, and jasmine blooms combined, these qualities worked incredibly well. The research revealed accuracy differences with moisture levels. The encouraging results suggest that a reliable machine vision system for classifying and recognizing agricultural and horticultural products may be possible.

Vera Schroeder et al. studied classifying complex odours using a chemoreceptive sensor array and a two-step machine-learning technique [55]. They identified optimal subsets of sensors for each Odor category, validated them with an independent dataset, and determined selectors through classification accuracy assessment and a comprehensive four-selector combination scan. Sample classification using k-nearest neighbours and random forest models achieved high accuracy for cheese and liquor samples in independent test sets. The methodology provides a versatile object classification approach for various challenges. Its expansion to handle a variety of gas sensing issues, such as speeding selection recognition from a wide range of molecules, detecting hazards, and authenticating food, still needs to be investigated.

Magnus et al. [56] used machine learning and optical spectroscopy to assess the safety and quality of food. They minimized the necessary wavelengths using feature selection, combining reflection and fluorescence spectroscopy, and utilizing a cascade classification. The ideal cascade for the method was a first classifier that used an extreme learning machine for reflection measurements and a second classifier that used SVM for fluorescence measurements, which was demonstrated using a case study on walnut processing. This work surpassed the cutting-edge in the culinary processing business with its excellent nut categorization, which concurrently improved food quality and safety evaluation and detected foreign items. The findings suggest machine learning's potential in multipurpose food processing applications, promising improved food safety practices.

A method for combining radio frequency identification (RFID) with artificial intelligence to identify destruction in a variety of foods and beverages, including liqueurs, alcoholic beverages, and infant formula milk, was presented by Muhammad Ali Imran et al. [57]. Ultra-high-frequency (UHF) sticker-type inkjet printer RFID tags were used to identify the levels of known contaminants in both infected and uncontaminated food products. The RFID tag on the meal products received a signal strength indicator (RSSI), and the message reflected phase was assessed through a

Tagformance Professional configuration. This accuracy of the sensing was improved by using the XGBoost machine-learning method. Utilizing RFID and machine learning technology, this research paves the way for widespread contamination sensing and gives consumers information about the safety and health of their food.

A semi-supervised generative adversarial network (GAN) technique built on deep CNN architecture was presented by Bappaditya Mandal et al. [58] to address the limitations of the food dataset identification process, such as the lack of labelled images. Trials using partially labelled data from the ETH Food-101 and Indian Food datasets showed that our strategy consistently outperformed previous approaches at different rankings. It outperformed cutting-edge methods despite having little tagged information. The study recognized the difficulties associated with stability and convergence during training, but it also emphasized the potential of GANs to improve food identification accuracy. Notably, the research did not explore strategies for improving recognition accuracy through a more robust GAN architecture to reduce reliance on labelled training data.

Based on the faster R-CNN network, Pengcheng Wei et al. created a system for extracting and classifying food images [59]. Using food image sets from the visual gene database; the scientists enhanced the faster R-CNN network to guarantee accurate food area detection. Tests on the food dataset Dish-233, which includes 49,168 images and 233 dishes, showed better results in image categorization than other techniques. The method also showcased excellence in image retrieval, indicating its efficacy in food image retrieval and classification tasks. However, scalability was not explored with larger datasets like Food-101. The research did not delve into retrieval, classification, or caloric estimation for food areas.

A smartphone-based system was created by Jayakanth Kunhoth et al. [60] to help visually challenged kids identify fruits and food dishes. In real-time photos, the system identified food items using a trained deep CNN model. They showed how to collaborate to merge multiple CNN architectures to produce a deep CNN model for meal detection. The computer had been trained using a custom set of 29 berries and food items. According to performance study, the ensemble model performed better in the custom dataset than the most advanced CNN models. However, the smartphone app did not address the CNN model's capacity to identify and recognize multiple food items simultaneously, nor did it support various platforms.

Huiru Zheng et al. [61] extracted features from various food picture records, including Food 5K, Food-11, RawFooT-DB, and Food-101, using initialized ResNet-152 and GoogleNet CNNs. Machine learning classifiers like artificial neural networks (ANN), support vector machines (SVM) using RBF kernels, Random Forest, fully connected neural networks, and Naive Bayes are trained using deep features that are retrieved from these CNNs. ResNet-152 deep features, especially in conjunction with SVM and RBF kernels, effectively-identified food items in the Food-5K dataset.

Table 2.1: Summary of literature survey of food recognition using Machine learning

Ref. No	Proposal	Technique	Inference	Research gap
[53]	Using machine learning, food recognition, and visibility, a novel algorithm is presented for weather visibility and food recognition.	CNN	High classification and accuracy.	Limited data set, dependency on web data.
[54]	Using machine vision, identify and categorize food grains, fruits, and flowers.	BPNN, quadratic discriminant function, linear discriminant function.	Robust system	Moisture sensitivity
[55]	Food Classification using Machine Learning And Chemiresistive Sensor Array.	Two-step machine learning approach, k-nearest neighbours and random forest models.	Versatile Classification.	Limited Extension.
[56]	Enhancing food classification through the use of machine learning and optical spectroscopy (combined).	Cascade classification, Extreme Learning, SVM, optical spectroscopy.	Effective Food Quality Evaluation	Limited Scope (Focused on Nut Classification)
[57]	Food contamination, Internet of Things and RFID detection, machine learning-assisted system.	RFID, machine learning, UHF RFID tags, XGBoost machine learning algorithm.	Enhancing & sensing accuracy	High cost
[58]	Partially Labeled Data-based Deep Convolutional Generative Adversarial Network Food Recognition.	semi-supervised GAN, deep CNN.	Better performance	Stability and convergence challenges
[59]	Classification and retrieval of food images using machine learning and visual cues.	Faster R-CNN network.	Better performance	Limited scalability

Ref. No	Proposal	Technique	Inference	Research gap
[60]	A smartphone-based food recognition system that makes use of many deep CNN models.	Deep CNN, ensemble learning.	Better performance, high accuracy.	Simultaneous multiple food item recognition not possible.
[61]	Different Food Image Datasets are Classified Using Supervised Machine Learning Algorithms and Deep Residual Network Features	Pretrained ResNet-152, GoogleNet CNNs, SVM with RBF kernel, ANN, RF, fully connected neural network, naive bayes.	Efficiency	Limitation in larger and uncontrolled datasets

2.3 FOOD RECOGNITION USING DEEP LEARNING

Neural networks were used by Ying Wang et al. [62] to build a system for food image identification. They created two versions of the two-stage learning technique, YOLO-SiamV1 and YOLO-SiamV2, by combining Tiny-YOLO with a twin network. Experiments validated the approach, demonstrating high overall identification accuracy without requiring human marking. It holds promise for practical use and widespread application. They also presented a technique that uses threshold segmentation technology to successfully separate foreign bodies from meals by detecting and identifying them. The experiment's outcomes showed how well the procedure worked to separate the desiccants from other materials and provide the intended result. This research highlights the potential for improving food safety through accurate foreign body detection in practical settings. The limitation lies in the inability to distinguish desiccants from foreign bodies solely through threshold segmentation.

Parisa Pouladzade et al. [63] developed a DL technique for accurately identifying various eating products in mobile-captured images to estimate meal calories and nutrition. To improve speed and precision, users quickly outlined the food area through touch input, employing image processing and computational intelligence for recognition, with training using single-item food images. During training, region proposal algorithms generated candidate regions, and submodular optimization aided positive region selection per food category. Testing involved generating candidate regions, computing classification scores based on CNN features, and offloading computational tasks to the cloud for prompt responses. Despite challenges like appearance variation, the system proved a robust tool for achieving high accuracy in food recognition, with occasional inaccuracies noted in test images.

Using the EfficientDet DL model, Prasan Kumar Sahoo et al. [64] created an AI-driven approach for identifying numerous dishes. The model addressed Taiwanese cuisine, including single-dish, mixed-dish, and multi-dish meals, in an effort to increase the efficiency of one-time meal reporting. This study demonstrated how well the model could realistically support food consumption, particularly when identifying a variety of regional specialties. The approach demonstrated high recognition performance for single- and mixed-dish food types, presenting a promising solution for streamlined dish reporting. Bounding boxes were used for each food item in the created automatic food identification model, which improved the user experience and decreased reporting time. This did not involve improving algorithm performance and integrating the technology into a real cloud computing and mobile platform in order to boost the accuracy of food intake reporting.

Barbara Koroušić Seljak et al. [65] concentrated on advancing dietary assessment and food-choice studies. Their integration of the well-known 'fake food buffet' approach with cutting-edge meal-matching technology made a substantial contribution. Their technique demonstrated efficiency in data gathering by using DL for fake-food image identification, which incorporated a unified network for image segmentation and pixel-level classification. This approach surpassed existing methods in pixel accuracy and laid the foundation for automating dietary assessment. Their work served as crucial groundwork in the field, paving the way for improved methodologies for comprehending and evaluating food choices. The sole component of the automated dietary evaluation procedure that is presently lacking was included in this research: an extension of the approach using a technology calculating meal volume or a tool automatically measuring weight (e.g., Foodscape Lab).

NutriNet is a newly defined deep CNN architecture used in an approach to food and drink image identification and recognition by Simon Mezgec et al. [66]. Using a dataset that included images taken with a smartphone camera and images from people with Parkinson's disease, they ran a real-world experiment. For real-world images, the top-five accuracy showed encouraging results. Additionally, NutriNet outperformed the baseline identification result when evaluated on the meal image collection data from the University of Milano-Bicocca 2016 (UNIMIB 2016). An online training component was introduced in order to enhance the food and drink identification model on fresh photos. In actuality, the model served as a part of a smartphone application that evaluated the meals of Parkinson's disease patients. Concerns about potential overfitting, arising from the continuous addition of food and drink images by Parkinson's disease patients, were not specifically addressed.

Gianluigi Ciocca et al. [67] provided a special dataset with 20 different food kinds in a variety of states, from solid to creamy paste, for food and state recognition. Using well-known CNN architectures, they conducted experiments on three categorization tasks: food categories, food states, and a combination of both. Instead of using end-to-end methods for classification, they used deep features taken from CNNs and SVMs to address the prevalent problem of sparsely labeled data. Comparative analysis with hand-crafted features demonstrated the superior performance of deep features across all tasks. Furthermore, their evaluation of generalization using an external dataset affirmed the robustness of CNN-based features to unseen data, emphasizing resilience to intra-class visual variations and challenges posed by different scales and lighting conditions in food images.

Using DL, Chang Liu et al. [68] sought to create a workable food identification system for nutritional evaluation inside the edge computing service architecture. The study concentrated on real-time food recognition using the edge computing service architecture and innovative DL-based food image identification algorithms. Numerous trials using actual data from the real world were carried out. The results proved that it achieved three objectives: (1) exceeded earlier studies about accuracy in identifying foods; (2) reaction time reduction was on par with the least amount of earlier techniques; and (3) energy consumption reduction almost matching the lowest of the state-of-the-art. The study, however, had no effect on the system's (reaction time and energy usage) or algorithms' (accuracy of detection) performance.

A CNN-based approach for food recognition was introduced by Qian Yu et al. [69]. It focused on transferring knowledge and improving the overall framework using

Inception-ResNet and Inception V3 models. When the system was used on the Food-101 dataset, it produced remarkable identification outcomes. With revised weights, all layers' fine-tuning greatly improved identification accuracy. Notably accurate and exhibiting quicker convergence was Inception-ResNet.

Table 2.2: A summary of the literature on deep learning-based food recognition

Ref. No	Proposal	Technique	Inference	Research gap
[62]	DL-Based Food Image Recognition and Food Safety Detection Technique.	Neural networks, Tiny-YOLO, twin network, threshold segmentation technology.	High accuracy and effective.	Limitation in Desiccant Differentiation
[63]	Mobile Multi-Food Recognition Using DL	DL method, image processing techniques, region proposal algorithm.	User friendly	Appearance and variation may impact the system's reliability
[64]	Multi-Dish Food Recognition Application Supported by Deep Learning for Dietary Intake Reporting.	AI-driven model, EfficientDet DL model.	Enhance user experience	Low accuracy
[65]	Automatic dietary assessment for the detection and standardization of fake food photos with a hybrid deep learning and natural language processing technique.	DL, unified network, pixel level classification,	Better performance, efficiency	Lack of weight measurement, and food volume estimation
[66]	NutriNet: A Dietary Assessment Tool Using DL Food and Drink Image Recognition	NutriNet, deep CNN	Better performance	Potential overfitting issue
[67]	Using Deep Features for Food Image State Recognition	CNN, SVM	Robustness	Computational complexity
[68]	A Novel Dietary Assessment System Utilizing DL-based Food Recognition on an Edge Computing Service Infrastructure	DL based algorithms, edge computing service paradigm	Better performance	Limited algorithmic improvement, low accuracy
[69]	DL Based Food Recognition	CNN, Inception-ResNet and Inception V3 model	Better performance, high accuracy	Untuned architecture, mobile implementation lacking.

2.4 FOOD CALORIE ESTIMATION

A hybrid architecture using DL algorithms was first presented by Tanmay Sarka et al. [70] to forecast the quantity of energy in each meal product in the serving dish. The system's three primary components were volume and calorie computation, segmentation, and classification. Using Mask (R-CNN), images were initially segmented, and YOLO V5 framework was then utilized for feature extraction and food item categorization. Identifying items facilitated the determination of dimensions, crucial for calculating food quantity, and subsequently, calories were computed based on the item's volume. Although the algorithms were successfully trained on dataset images to produce accurate calorie estimates, difficulties continued since food comes in various forms, sizes, and presentations. Notably, creating accurate ground truth data for training was expensive and time-consuming.

Using a handheld camera, Razib Hayat Khan et al. [71] developed a system to recognize images of routine meals using parameter-optimized CNN. After completing the food item identification procedure, the related nutritional information and calories were computed using pre-existing knowledge about the food class. The method proved easily trainable and applicable to customized datasets with heightened accuracy through simple linear operations. The study showed the excellent accuracy of the technique and its potential to automate and simplify current manual calorie estimating processes into a real-time process. A more thorough understanding, however, would have required training the model with a variety of food images in order to facilitate the identification of a wide range of food items. This constraint was ascribed to the dearth of a high-quality image dataset that satisfied the requirements.

Using Mask R-CNN, Kidong Lee et al. [72] presented an actual time-vision-based method for calculating energy and dividing solid-volume meal instances. Model training was done with RGB images, and testing was done with a basic monocular camera. Gimbap served as a demonstrative solid-volume food example. In order to improve identification accuracy based on Gimbap posture in plates, the study first established a labelling technique for Gimbap image datasets. After then, an enhanced model for gimbap identification was created by refining the architecture of the Mask R-CNN. A calorie estimation approach was formulated by merging calibration and

Gimbap instance segmentation results. This solution, which used mask information to account for unseen food, showed improved accuracy in food segmentation and calorie prediction compared to faster R-CNN-based algorithms. The effectiveness of these techniques was confirmed while considering the uncertainties resulting from calibration and detection mistakes caused by different illumination and image quality.

A neural network-based model was created by Shahriar Ahmed Ayon et al. [73] to identify food items from provided images and provide the estimated calorie value of the identified food items. They compiled a dataset of approximately 23,000 images covering 23 different food categories. Their system detected various cuisines by utilizing Fusion V3's capabilities to teach a CNN. Achieving accuracy with this model, they deployed it on a webpage where users could upload food images to predict food items and their estimated calories. The system demonstrated high efficiency in object or image classification. However, it typically performed poorly with images captured from a top view.

Using DL, Takumi EGE et al. [74] concurrently learned food calories, categories, components, and cooking instructions to create a system for predicting food calories from food images. They anticipated that training these aspects simultaneously would improve performance compared to training them independently. For this purpose, they utilized a multi-task CNN. They also produced two datasets, one from an American recipe website and the other with calorie annotations from Japanese food websites. They trained single-task and multi-task CNNs in their studies and compared their performance. The outcomes demonstrated that the multi-task CNN performed better than the single-task CNNs in terms of food category and calorie estimation. They did not, however, create a smartphone application that combines food dish identification with a real-time food calorie estimate.

S Jasmine Minija et.al [75] introduced a dietary assessment system that combined segmentation and classification techniques. A specific kernel function was used in the segmentation procedure, which split the image using the CSW-WLIFC methodology. At this point, the segmented sections' shade, structure, and symmetry were recovered. The Whale Levenberg Marquardt Neural Network (WLM-NN) classifier had to be taught to recognize various items in the dish image. A calorie calculator was also created to determine how much energy is in every meal. The simulation outcomes showed that this model performed better than previous techniques.

The goal of Shaikh Mohd. Wasif et al. [76] was to create software that uses machine learning and image processing to give users information about the calories in meals. Two images, side and top, were needed, together with a probing object—like a coin—for volume reference. Their process involved object detection using Faster R-CNN, image segmentation via the grab cut algorithm, volume calculation of food items, and subsequent calorie estimation. By leveraging Faster R-CNN, they achieved real-time object detection, bypassing selective search for region proposals. The grab cut algorithm facilitated foreground extraction with minimal user intervention. Their implementation successfully determined food volume, mass, and calorie content. However, challenges may arise regarding accuracy, particularly with complex food shapes or variable lighting conditions.

Mrs. Srilatha Puli et.al [77] developed a model to measure calorie intake using DL algorithms. They employed CNN, Random Forest, and SVM algorithms for food recognition and detection, achieving improved accuracy. Utilizing a publicly available food image dataset, CNN was utilized for image recognition, and models were trained with dataset information. Further accuracy enhancements were made through optimization and hyperparameter tuning. They developed a function to determine calories based on detected fruits, considering their average calorie values. The CNN model demonstrated higher accuracy, as reduced volume error estimation indicates. However, the paper does not include calorie estimation outcomes. The model did not expand the dataset or address imbalanced data distribution, potentially increasing errors.

Using multi-spectral images, Ki-Seung Lee et al. [78] created a method for automatically identifying the type of food and its caloric content. To categorize food products and calculate their calorie counts, they used CNN. 10,909 images representing 101 different food varieties made up the training data. Based on various wavelength values, they looked at objective functions such as classification accuracy and mean absolute percentage error (MAPE). Piecewise selection was used to find the best wavelength combinations. Validation tests were performed on 3636 training-set images of various food categories. UV/VIS/NIR band light sources in conjunction with RGB images greatly increased performance in food item categorization and calorie estimation compared to earlier approaches. It did not investigate the relationship between ingredient quantification and multi-spectral imaging methods.

Sanghyun Seo et.al [79] utilized a SVM and multilayer perceptron (MLP) model to identify food types and estimate their calorific values. The study employed a variety of preprocessing methods, segmentation, and feature extraction, with an emphasis on specific food items. Following extraction, the features were fed into SVM and MLP classifiers to enable recognition. The program automatically computes food item calorie values based on volume to help diabetes patients maintain a healthy diet and BMI. The findings showed that food item identification and calorific content estimation accuracy were acceptable. In order to monitor food consumption and avoid illnesses, this work did not include analyzing complicated food items using optimum feature extraction techniques and integrating multiple machine learning models, clustering approaches, and external devices.

DL approaches for food recognition and calorie calculation were investigated by Yuan Gao et al. [80], addressing the real-world requirements of diabetics controlling their blood glucose levels at home. Their method consisted of employing images or level sensor reconstructions to determine a quantity of meals, followed by DL to calculate the weight and calories. They created a technique where users would take an image of the meal, and an object identification algorithm would determine its category and location. The food's weight was determined via a weight prediction algorithm, making it possible to calculate calories. Their approach helped with calorie management by allowing users to regulate meal portions, making it appropriate for home cooking. Nevertheless, because it did not concentrate on assembling more training sets and refining model structures to increase accuracy and speed, using a white plate may have restricted its efficacy.

Table 2.3: Summary of literature survey of calorie estimation

Ref. No	Proposal	Technique	Inference	Research gap
[70]	Hybrid DL Algorithm-Based Approach for Calorie Estimation and Food Recognition	DL, mask (R-CNN), YOLO V5 framework	Accurate calorie prediction	Time consuming, expensive
[71]	CNN-Based Real-Time Food Calorie Estimation from Images that is Simple and Parameter-Optimized	Parameter-optimized Convolution Neural Networks	Better performance	Low accuracy
[72]	Using Mask R-CNN for Calorie Estimation and Solid-Volume Food Instance Segmentation	Mask R-CNN, labelling approach	High accuracy	Uncertainties in detection and calibration
[73]	FoodieCal: A CNN-Powered Calorie Estimation and Food Detection System	Neural network, CNN, Inception V3	High efficiency in classification	Poor performance with top-view images.
[74]	Food Calorie Estimation Based on Image and Recipe Data	DL, multi-task CNN	Better performance	No real-time calorie estimation capability.
[75]	For daily nutrition assessment, fuzzy clustering and a whale-based neural network are used to identify foods and estimate calories.	CSW-WLIFC model, WLM-NN classifier, kernel function	Better performance	Complexity for inexperienced users.
[76]	Estimating food calories with machine learning and visual processing	Faster R-CNN algorithm, grab cut algorithm	Effective	Low accuracy
[77]	CNN Food Calorie Estimation	DL, CNN, Random Forest, and SVM algorithms	High accuracy	Imbalanced data distribution may lead to errors.
[78]	Using CNNs for Multispectral Food Classification and Caloric Estimation	CNN	Improved performance	Limited exploration of ingredient quantification.
[79]	Using neural networks for food type recognition and calorie estimate	SVM and MLP model	Accurate calorie estimation	Limited to single food items.
[80]	CEP: Calculating Calorie Content from Food Photos	DL techniques	Enables home users to manage calorie intake.	Low accuracy, high processing time

2.5 FOOD DEFECT DETECTION

Using a UNet4+ segmentation model, Tiezhu Shi et al. [81] looked at the detection of mustard pickles' cerebral nerves. They categorized its pickled's edible portion mustard and its backdrop as different classes, and cortical fibers as the target class. They combined MS with HD images to enhance image quality by employing an improved imaging fusion technique. Preprocessing methods were applied to each image, resulting in 150 training and 50 test images. Next, cortical fiber detection was trained into the UNet4+ model. Successful detection was demonstrated on MS, HD, and fusion images, spanning all image formats. The comparison between the UNet++ and UNet3+ models revealed better results. Regarding accuracy, there was still room for development. To increase inference speed and decrease parameters, they did not increase the dataset's size or prioritize creating more effective DL models.

Alessandro Ulrici et al. [82] used colorgrams to detect red skin defects in raw hams, aiming for more objective evaluations. They selected 95 raw ham samples and converted their RGB images into colorgrams. These were utilized in conjunction with a wavelet packet transform-based method (WPTER) and partial least squares discriminant analysis (PLS-DA) to generate categorizing models. With only three variables, feature selection allowed for the accurate forecast using a different sample and allowed for the distinction of faulty samples. The reconstructed images using selected features validated the classification model's reliability. This study demonstrated how crucial appropriate feature selection algorithms are for finding useful features, cutting down on computing time, and supporting critical analysis of algorithmic decisions.

Deep CNN (DCNN) models were employed by Zhifeng Xiao et al. [83] to identify surface flaws in potatoes. By modifying a basic technique using various distinct DCNN models—SSD Inception V2, RFCN ResNet101, and quicker (R-CNN) ResNet101—they utilized transfer learning. After training on their dataset, they achieved a noteworthy level of accuracy, with RFCN ResNet101 showing the top results in terms of speed and accuracy. Its capacity to generalize outside of the training set was demonstrated when it was chosen as the final model to be tested on fresh data. However, the researchers did not explore data augmentation due to time and cost constraints, which could have enhanced the training set's diversity and size. Despite this, data augmentation has been proven effective in other studies, suggesting its potential for further improving automated potato defect inspection systems.

For the purpose of fruit defect identification, Van Huy Pham et al. [84] created a hybrid image splitting method built around the separation and combine method. The original image was first segmented into areas built on linear shading in the L*a*b* area using the k-means technique, which led to over-segmentation. Next, a combined process utilizing a minimal span structure was used to repeatedly integrate related sections into more homogeneous ones using a graph representation.

Arthur Z. da Costa et al. [85] made a significant contribution to the field of exterior defect identification using DL. They created an unedited dataset with 43,843 images that included flaws from the outside. Deep residual neural network (ResNet) analyzers were developed to detect external problems through feature extraction and fine-tuning. This optimal model represented a significant achievement in external defect detection using computer vision, demonstrating the practical application of their research.

Raheel Siddiqi et al. [86] developed two distinct convolutional object detection algorithms that can complete the apple defect detection challenge. The SSD and the YOLOv2 object detection frameworks were the foundation for the new systems. A dataset including 244 images of defective apples was developed to train and evaluate the detection algorithms. The system's performance metrics were positive. On the other hand, compared to the YOLOv2-based system, the SSD-based system performed significantly better.

Using a computer vision system, Chunjiang Zhao et al. [87] created an automated approach to identifying faulty apples. They corrected uneven lighting and counted defect candidates, like stems and calyces. Based on candidate counts, apples were categorized as sound (no defects) or defective. They analyzed features like color and texture for uncertain cases, using the I-RELIEF algorithm to find relevant features and weights. Then, a weighted relevance vector machine classifier determined if candidates were true defects or stems/calyces. Finally, apples were classified as sound or defective based on candidate categorization. This method effectively detected defective apples but had limitations, like only considering one view during inspection.

Table 2.4: Summary of literature survey of food defect detection

Ref. No	Proposal	Technique	Inference	Research gap
[81]	Using a Recently Developed DL Model in conjunction with Multispectral and High-Definition Imaging to Detect Food Defects	UNet4+ segmentation model	Better performance	Low accuracy, high processing time
[82]	RGB imaging-based automated food defect identification and visualization: application to raw ham red skin defect detection	PLS-DA, WPTER	Effective, low computation time	Colorgram conversion error possibility.
[83]	Detection of Potato Surface Defects Through Deep Transfer Learning	DCNN model, RFCN ResNet101, transfer learning, SSD Inception V2, and Faster (R-CNN) ResNet101	Accurate defect detection	Failed to diversify dataset, limiting system's potential.
[84]	An image segmentation method utilizing a graph-based algorithm and k-means clustering for fruit defect identification	k-means algorithm, minimum spanning tree, Euclidean color distance measurement	Efficient computation	No improvement in merging conditions
[85]	DL-based computer vision-based exterior flaw detection on tomatoes	DL, Deep residual neural network	Better performance	Requirement for extensive labelled data.
[86]	Using cutting edge object detection methods, automated apple fault detection	SSD and YOLOv2 framework	Better performance	Limited dataset
[87]	Using a weighted RVM classifier and automated lightness adjustment, computer vision can identify faulty apples.	I-RELIEF algorithm, RVM classifier	Effective detection	Single-view inspection leads to incomplete assessment.
[88]	Food Mem: Accurate and Almost Real-time Food Video Splitting.	Segmentation Trans former (SETR)	Higher performance, with 2.5% greater mAP and 58 times faster processing than current models.	segmentation models, such as flickering and prohibitive inference speeds in video processing
[89]	Seamless Segmentation with Occlusion Awareness.	Occlusion-Aware Seamless Segmentation (OASS)	SynPASS and DenseP ASS, obtains 45.34% and 48.08% in mIoU	OASS addresses object occlusion and FoV occlusion limitations in segmen tation tasks.

2.6 SUMMARY OF LITERATURE REVIEW

The chapter extensively examines Deep Learning and machine learning techniques for food identification and calorie estimation. It looks at many methods for spotting food and figuring out how many calories it has, as well as other methods for identifying food defects. Despite the capabilities of these techniques, they exhibit limitations. This research introduces an innovative approach for food detection and calorie estimation to address these challenges.

Food detection using deep learning has seen significant advancements, primarily driven by convolutional neural networks (CNNs) and their variants, such as Faster R-CNN, YOLO, and SSD, which have improved the accuracy and efficiency of food recognition tasks. These models benefit from large food image datasets like Food-101 and UECFOOD-256, though challenges persist in intra-class variability, inter-class similarity, and complex backgrounds. Transfer learning has been widely adopted to overcome data limitations, while real-time detection for mobile applications has gained traction. Current research is exploring multimodal learning, 3D food recognition, and explainable AI to further enhance detection accuracy and applicability in areas like dietary monitoring, restaurant automation, and health tracking.

Volume estimation using deep learning focuses on accurately determining the volume of objects, particularly in applications such as medical imaging, food portion estimation, and industrial measurements. Deep learning models, especially convolutional neural networks (CNNs), combined with 3D imaging techniques like depth sensors, stereo vision, and point clouds, have significantly improved volume estimation accuracy. Approaches such as volumetric CNNs, 3D U-Net, and neural networks incorporating geometric cues are commonly used. Challenges include handling occlusions, variations in object shapes, and ensuring robustness in real-world environments.

CHAPTER 3

PROPOSED METHODOLOGY

3.1 INTRODUCTION

This section presents the methodology for developing a Real-Time Calorie Assessment in Food using deep Learning CNN. It introduces the proposed model, the algorithm and techniques used, and the flow of the model framework. More importantly, it outlines the practical steps involved in developing a real-time calorie assessment, demonstrating the potential real-world application of the research.

3.2 PROPOSED METHODOLOGY

The proposed model is divided into three phases in which first phase is to create custom dataset. In second phase dataset has been extended and detect the food items using various YOLO algorithms and in final phase, a hybrid architecture is developed to assess the calorie content of different food items within a lunch bowl by utilizing deep neural networks. Three essential parts make up this architecture: volume and calorie computation, segmentation, and classification. Figure.3.1 shows the Implementation of the deep learning algorithm in the Android application.

1) Capture of Image

2) Image Segmentation

3)Recognition of food

4) Volume of food

5) Estimation of Calorie

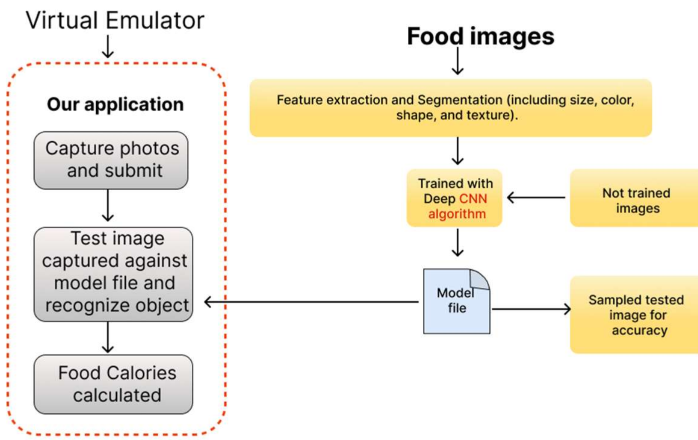


Figure 3.1: Proposed Methodology

The intriguing fields of DL for real-time calorie estimation and food identification are investigated in this paper. The primary objective is to address the shortcomings of the current calorie calculation method, including the dataset, meal volume, and distance measurement difficulties. The proposed work introduces innovative concepts for real time calorie measurement. The primary contributions are to prevent the above-mentioned problems and ensure the food recognition and calorie estimation system is strong and reliable. Figure 3.2 shows the steps of calorie estimation using deep learning.

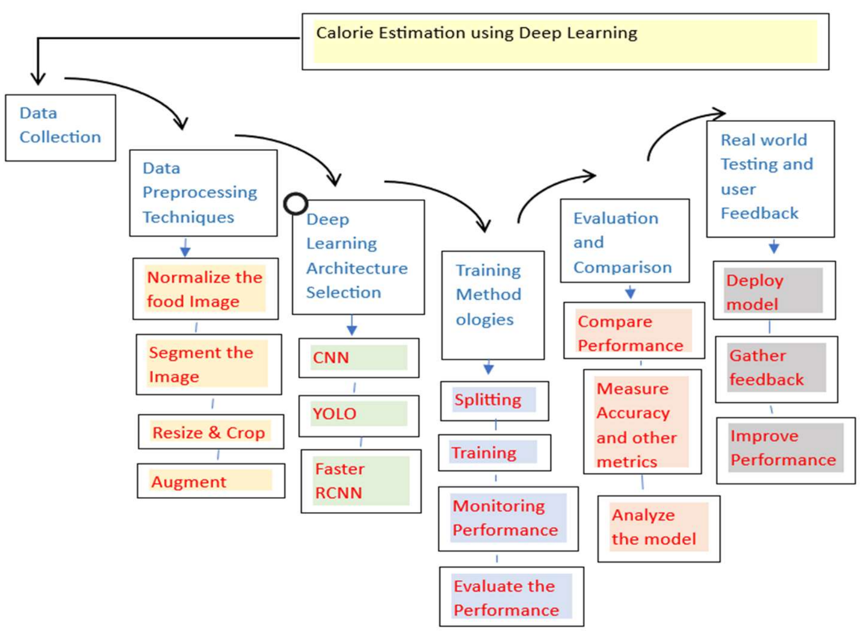


Figure 3.2: Calorie estimation using Deep learning

3.3 PHASE 1

Indian cuisine is well liked all around the world. Rapid and accurate item identification techniques are becoming increasingly in demand due to rise of facial recognition, driverless cars, intelligent video surveillance, and many more human counting applications [90].

These techniques include localizing an image by using the best bounding box to be draw n around each object and identifying and categorizing each one. One of the most common duties in visual computing is object detection in images. The use of computing vision, which specializes in classifying and identifying pictures, videos, and even live streaming material, is called object detection. They are trained on abundant annotated images to enable object detection models to operate this system with fresh data. The explanation of computing vision tasks is as follows: The process of predicting a things class in an image is called image classification. Rather than using bounding boxes to identify objects, assembling pixels with comparable characteristics is called integration. [91]. While object detection recognizes all things and their boundaries without regard to location, things localization ambition to find one or more products inside an image [92]. To maximize the detection accuracy of these objects, these object identification models are trained using a large amount of visual input. DL models to detect Indian food dishes can greatly improve computer vision applications in the food industry. However, there aren't enough datasets for training these models specifically on Indian food. The IndianFood-7 Dataset, which consists of over 800 images with over 1700 comments on seven different Indian cuisine items, is presented as a solution to this problem. Several image-augmentation approaches were used in this study to strengthen the model. Using F1 score, mAP, accuracy, and recall, the study compares multiple state-of-the-art object identification model versions, including YOLOR and YOLOv5[93,94]. The effectiveness of these models was evaluated by looking through and analysing the models' predictions on images from the test dataset. In order to improve the effectiveness of computer vision-based techniques for identifying popular Indian cuisine items, this project aims to expand the Indian Dataset.

3.3.1 Methodology

The IndianFood-7 dataset is trained on a DL-based object identification method. This chapter focuses on identifying seven common Indian foods in images. The advanced object detection models YOLOv5 and YOLOR were used in different versions. Metrics, including F1 Score, mAP, precision, and recall, were used to evaluate and compare the models' performances [95].

3.3.2 Data Preparation

First, a list of well-known Indian dishes is created. The dataset is then created by gathering a diverse range of images of Indian food from various classifications based on this list. The locations of each food item inside the images are indicated by annotations on the bounding boxes. The final collection, called IndianFood-7, is a rich and diverse resource, with more than 1700 annotations spread over many classes of common Indian meal products, totalling over 800 images. Visit this link to learn more about the dataset and the quantity of annotations for each class: https://universe. roboflow.com/indianfood/indian_food-pwzlc. The labels for the Indian cuisine classes and the quantity of inscriptions in the dataset that correspond to them are shown in Table 3.1.

Table 3.1: Displaying the labels for food classes and the quantity of annotations for each category found in the IndianFood-7 dataset.

S.No.	Classes of Indian Food	Annotations in number
1	GulabJamun	250
2	Samosa	382
3	Poha	206
4	Idli	309
5	PalakPaneer	193
6	Besan Cheela	215
7	Dosa	198

The Indian Food-7 dataset is a specialized collection of images aimed at facilitating the recognition and classification of Indian food dishes. These annotated images from the Indian Food-7 dataset demonstrate the variety of Indian cuisine and the complexity involved in accurately recognizing and classifying these dishes using machine learning models. The dataset is instrumental in advancing food recognition technology, particularly for Indian dishes. Figure 3.3 displays several examples of annotated images from the dataset of Indian Food-7.

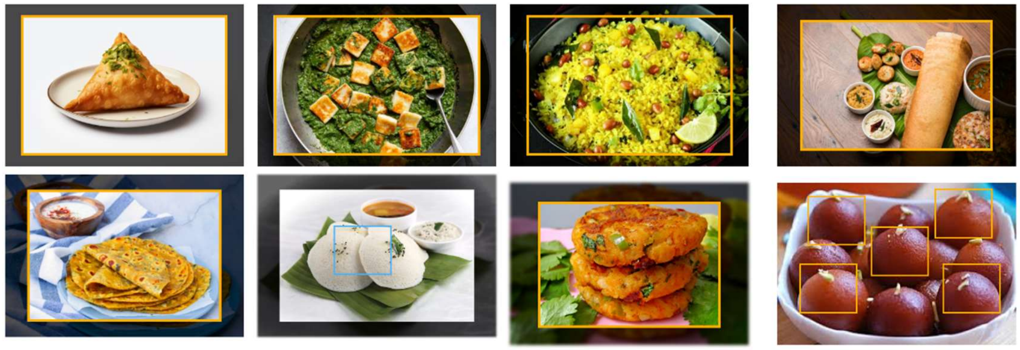


Figure 3.3: Samples of dataset (IndianFood-7) Annotated Images

To prepare the **IndianFood-7 dataset** for machine learning, follow several essential steps. These steps involve gathering, cleaning, preprocessing, and organizing the data so that it can be fed into machine learning models, specifically for tasks like food recognition. Algorithm1 explains the necessary steps in order to prepare the IndianFood-7 dataset for machine learning. It consists of the division of the dataset into training, validation, and testing sets, normalization of images, resizing of images, the addition of various rotations and flipping to the training data, and at last, returning the prepared data set for training as well as evaluation of the model. Figure 3.4 provides an illustration of the dataset preparation procedure. The raw dataset may contain inconsistencies, such as duplicate or irrelevant images, misclassified or mislabeled images, images with poor quality (blurry, low-resolution). To clean the data,remove duplicate images using a hashing method to compare image similarity and remove duplicates, eliminate images that are too blurry or have very low resolution, as they may negatively affect model performance, verify the correctness of the labels. Misclassified or poorly labeled images can introduce noise into your training process. In Data augmentation, ensure that all images are accurately labeled according to their respective food categories. The labels might already be provided in the dataset, but if not, need to manually label the images. Annotations can typically be done in a CSV or JSON format where each row links an image file to its category. To increase the diversity of the dataset, data augmentation techniques is applied which helps to simulate real-world variations, making model more robust and improving its ability to generalize. Split the dataset into three main parts as 70% for training, 15% for validation, 15% for testing. Preprocessing ensures that all images are in a uniform format and size, which is essential for training most machine learning models.

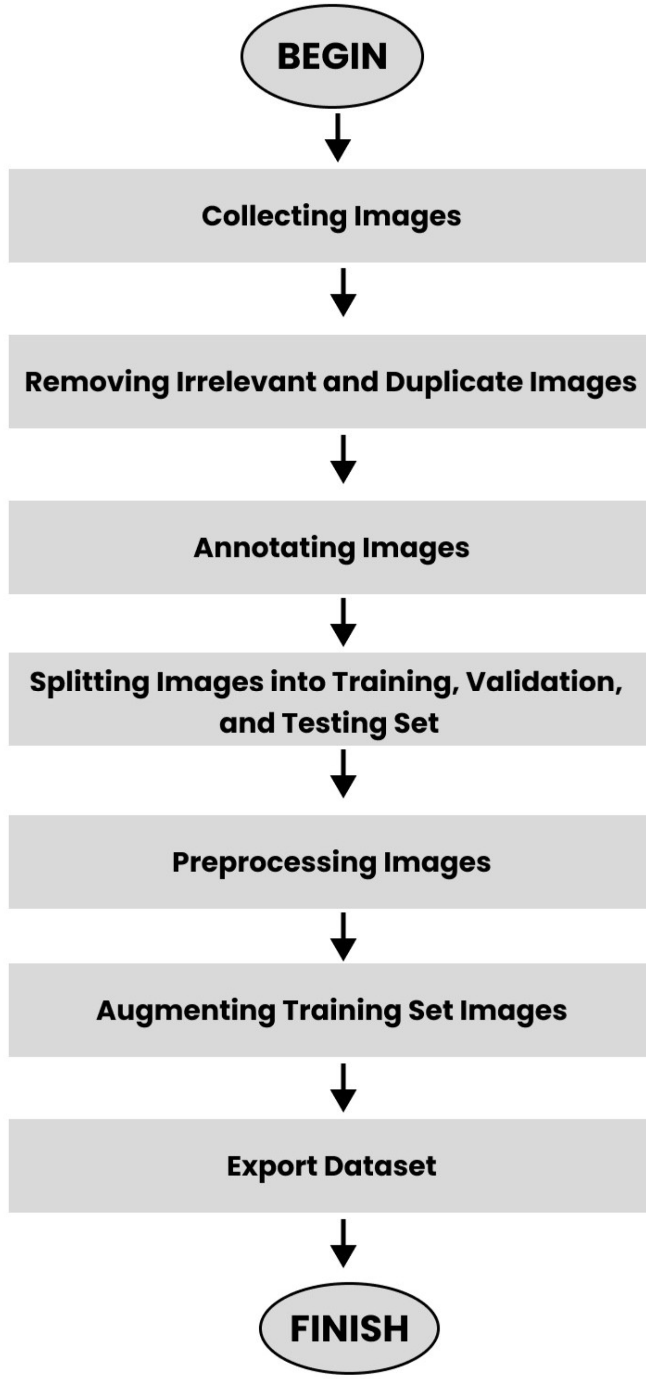


Figure 3.4: Flow of dataset preparation

Algorithm 1 IndianFood-7 Dataset Preparation

Input: Dataset $D = \{I_1, I_2, \ldots, I_m\}$ with annotations $A = \{A_1, A_2, \ldots, A_n\}$.
Output: Processed training, validation, and testing sets.
Step 1: Dataset Splitting
Split D into D_{train}, D_{val}, D_{test} such that:

$$D_{\text{train}} = 0.7D, \quad D_{\text{val}} = 0.15D, \quad D_{\text{test}} = 0.15D$$

Step 2: Preprocessing and Augmentation of $D_{\textbf{train}}$
for each I_i in D_{train} **do**
    Resize $I_i \to I'_i$ to a fixed size (H, W).
    Normalize pixel values: $I'_i = \frac{I'_i - \mu}{\sigma}$.
    Augment I'_i: Apply transformations $\mathcal{T}$ (rotation, flip, etc.) to generate $\mathcal{T}(I'_i)$.
end for
Step 3: Final Dataset
Return $\{D_{\text{train}}, D_{\text{val}}, D_{\text{test}}\}$ for model training and evaluation.
End.

3.3.3 Experimentation on object detection models

Using the internal dataset, a comparison and training were done on the YOLOR-P6, YOLOv5l, YOLOv5s, and YOLOv5m techniques due to their superior performance in terms of precision and quickness. The GPU system of the Nvidia Tesla T-4 was used to train the models over 150 epochs. Figure 3.5 provides an illustration of the experimental procedure.

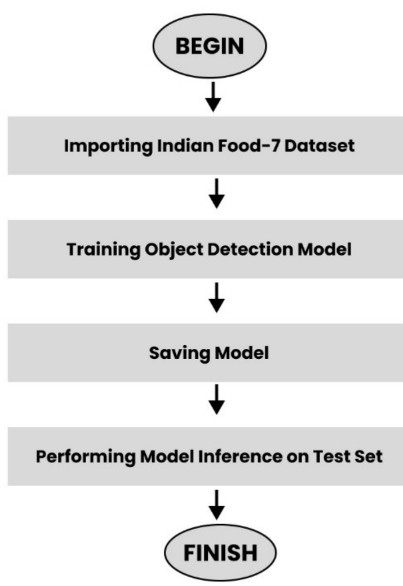


Figure 3.5: Flow of Experimentation

Algorithm 2 identifies the experimental procedure of training and comparing different models, namely MYOLOR-P6, MYOLOV51, MYOLOV5S, and MYOLOV5m, on the IndianFood-7. It includes training every one of these models on the training data set for some number of epochs, checking the accuracy and training time on the validation data set, and comparing the models and then choosing the best model. The experiments are performed on an Nvidia Tesla T-4 GPU.

Algorithm 2 Experimental Flow for Model Training and Comparison

---

Input: Training dataset D_{train}, validation dataset D_{val}, Models $\{M_{\text{YOLOR-P6}}, M_{\text{YOLOv5l}}, M_{\text{YOLOv5s}}, M_{\text{YOLOv5m}}\}$, Nvidia Tesla T-4 GPU.
Output: Performance comparison of models in terms of precision and speed.
Step 1: Training Setup
Set number of epochs $E = 150$.
Initialize GPU system (Nvidia Tesla T-4).
Step 2: Training Loop
for each model $M_i \in \{M_{\text{YOLOR-P6}}, M_{\text{YOLOv5l}}, M_{\text{YOLOv5s}}, M_{\text{YOLOv5m}}\}$ **do**
    Train M_i on D_{train} for E epochs.
    Validate on D_{val}.
    Record precision P_i and inference speed S_i.
end for
Step 3: Comparison
Compare models based on precision P_i and speed S_i for each M_i.
Select the best model M_{best} based on the highest P_i and lowest S_i.
End.

3.3.4 Summary

In order to improve computer vision-based methods' ability to recognize common Indian food items, this work presented the IndianFood-7 Dataset. The IndianFood-7 Dataset, which includes seven well-known Indian cuisines, was released. Various image augmentation methods were used to improve the model's resilience [96]. The goal was to train a Deep Learning-based detection system to identify Indian culinary items. Different versions of YOLOv5 and YOLOR models, renowned for image object detection, were utilized in the experiments. Model performance was evaluated through analysing and assessing their predictions on images from the test dataset.

3.4 PHASE 2

India is renowned for its rich cultural heritage and diverse cuisines, which are celebrated worldwide for their Flavors and tastes. In the modern world, following a person's prescribed diet is crucial [97]. Consuming too many calories can be harmful and cause a number of illnesses. The imbalance between the quantity of calories ingested and the calories it contains is the primary basis of fat. The pace of the increase in adult obesity is concerning. Diabetes is one of the major chronic diseases that might increase obesity. According to one type of study, women had a lifetime risk of 64.6% for diabetes, compared to 55.5% for males over the age of 20 [98–100]. The first step to encouraging good health is keeping track of your regular food consumption. Tracking nutritional scores can be facilitated by counting calories. As a result, it is widely accepted that eating too much food is the primary cause of many ailments in humans. Estimating food quantities in each dish poses significant challenges for individuals. Begin effortlessly counting the number of calories we consume each day. The dramatic shift in the generation's eating preferences toward healthier options highlights the need for easier and more precise verification. Our daily intake of energy from meals on a daily basis requires a tool that can precisely calculate the calories in this meal, in addition to being able to identify them with exceptional accuracy. Everyone has a functional portable gadget for this reason. Users may precisely determine food types and portion amounts with the aid of a computerized system. Comparing computerized technologies to older approaches reveals several advantages. However, the absence of a uniform dataset needed to assess DL-based item identification models has restricted the detection of Indian cuisines using computer vision techniques. Thus, systems that can offer data on serving sizes, calorie content, and consumption recommendations are required. The IndianFoodNet dataset, a comprehensive resource, has been made available to solve this problem. It includes more than 5,000 annotations and more than 5,500 high-quality images covering thirty distinct Indian culinary dishes, providing a robust foundation for research and application. The goal of this effort is to further uses of image processing that deal with identifying and detecting Indian cuisine. An analogy investigation of many cutting-edge object identification models, including YOLO5, YOLO7, and YOLO8, is also included in the study. By objectively examining the models' predictions on images that have been separated for testing within the dataset, the effectiveness of these models is assessed.

3.4.1 Methodology

The procedure for preparing the dataset is described in this section. Figures 3.6 and 3.7 show the general workflow of the dataset preparation process. The graphics show how the dataset preparation process is divided into two phases: • The dataset preparation procedure. •Training of an object detection model.

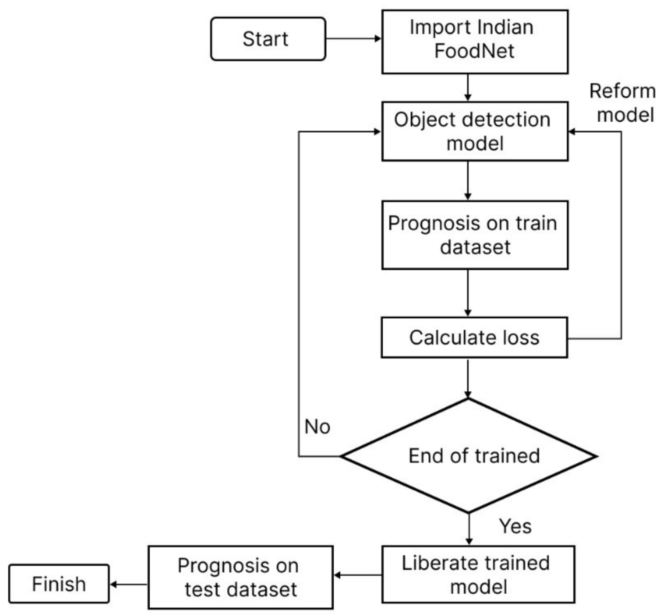


Figure 3.6: Flow of approach

3.4.2 Dataset Description

Earlier, Google and Bing were utilized to gather pertinent food images online. Bounding boxes containing the meal product dimensions are included. At that point, meal items could be prepared to practice object recognition.

As the last or second stage, model architectural creation was completed. By now, the relevant optimizers and the reduction in function were created, and the hyperparameters had been configured. The model could forecast the results on the training set because of all of these. It would minimize the value loss by updating itself in compliance with the specifications. Following this prescribed training, the model was able to make deductions from the test set. Using never-before-seen data, it demonstrated the trained model's effectiveness.

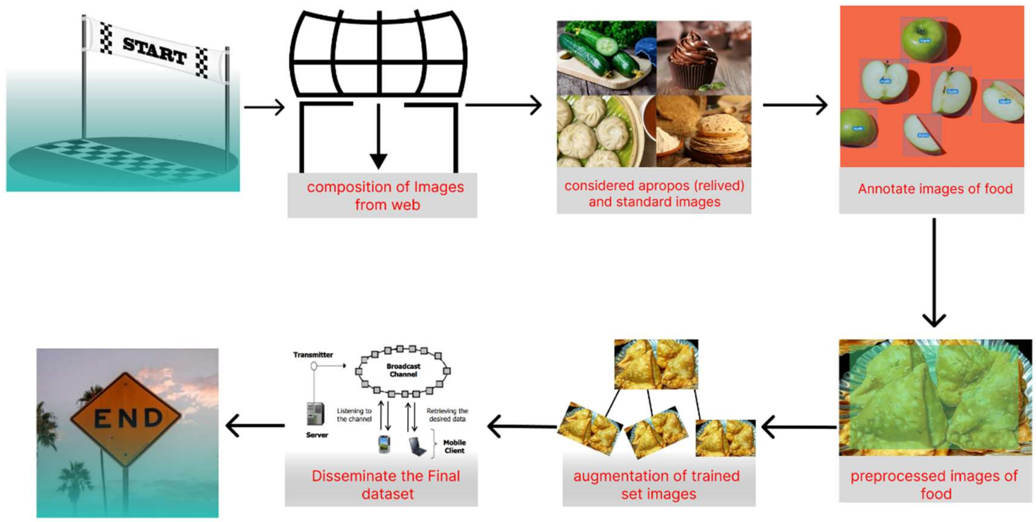


Figure 3.7: The Process of Preparing the Dataset

For the research, a dataset called IndianFoodNet was created. It consists of images of food products that might lead to food object recognition. It also includes an annotation box (or boxes) for each image, which makes it possible to find the exact location for the caloric thing.

The menu items were reduced to the following:

a) Meal products include dishes from different Indian regions, including those from the east, west, north, and south, specifically in Table 3.2.

b) Food items that are distinguishable solely by their visual attributes (this is due to the significant reliance on computer vision techniques, which were employed to determine the look of food items in any given image). It is the most significant criterion used to narrow down the meal options.

c) Finally, food products are regularly and consistently ingested.

Table 3.2: List of classes created for Dataset.

S. No.	Class Name of Food	S. No.	Class Name of Food
1	Bhatura	16	White Rice
2	Idli	17	Cheela
3	Gulab Jamun	18	Dum Aloo
4	Fish Curry	19	Ras Malai
5	Samosa	20	Green Chutney
6	Ghevar	21	Shahi Paneer
7	Rajma Curry	22	Jalebi
8	Dal	23	Kabeb
9	Aloo Gobi	24	Kheer
10	Aloo Masala	25	Kulfi
11	Dosa	26	Lassi
12	Bhindi Masala	27	Mutton Curry
13	Biryani	28	Onion Pakoda
14	Chai	29	Palak Paneer
15	Coconut Chutney	30	Poha

3.4.2.1 Collection

For crawling, Bing and Google were utilized. These used JavaScript queries to deliver the necessary, appropriate images. After being carefully examined, the images were blended. Images that had RGV channels that were not full might be deleted. These were eliminated from the dataset after being found via image hashing [101,102]. Several images that were out-of-date or irrelevant to the purpose of this research were discovered after searching the search engines.

3.4.2.2 Annotation

After being collected, every image was annotated on the Roboflow website. This was made possible by its find-noting tool, which also allowed annotation of the full dataset without requiring the download of any additional software.

The robot sources helped define the tool, which was quite useful in speeding up the annotating process overall. The trained model refreshes annotations as new images arrive in the dataset. Consequently, extensive manual testing was done on the annotations to guarantee their quality and correctness [103]. To get the dataset and learn more about how many comments are on every tag, click this link: https://universe.roboflow.com/indianfoodnet/indianfoodnet.

3.4.2.3 Pre-processing

The training of models on an initial set, and their performance is evaluated on the validation set. Throughout the training phase, the ideal model is kept, allowing inferences to be made about the test set. Figures 3.8, 3.9, and 3.10 display samples of training, valid, and experiment images, respectively. The accuracy, known, and mean average precision (mAP) of YOLO models are assessed across five training epochs in the evaluation metrics section. Based on the validation set, the bounding box regression loss decreases with training.

Conversely, the model's mean accuracy, recall, and precision all rose as it was trained, indicating that the model's prediction skills improved. Every model's training summary displays this trend. Following annotation of the images, the dataset is further split into a train-shaped dataset (70%), a test set (10%), and a validation subset (20%) of the dataset.

Figure 3.8: IndianFoodNet Dataset Sample of trained images.

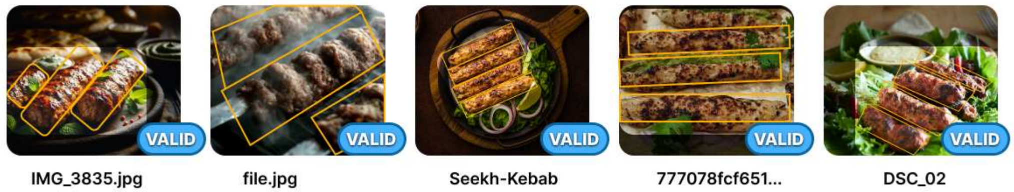


Figure 3.9: IndianFoodNet dataset Sample of valid images.

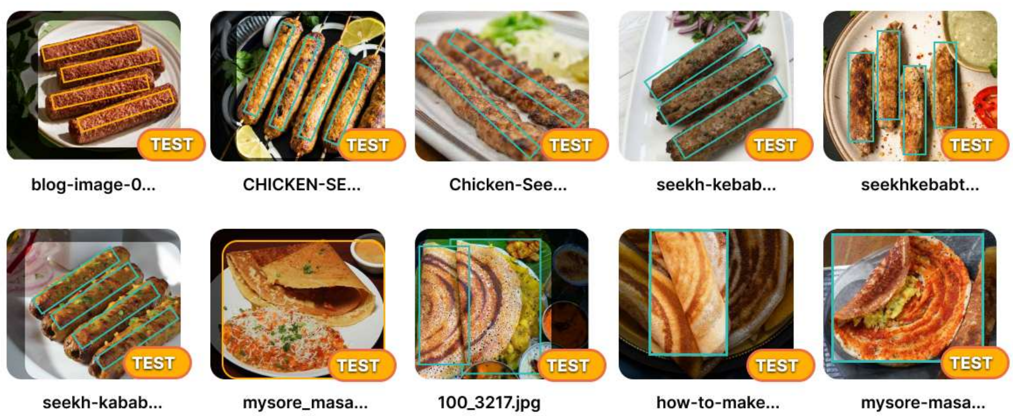


Figure 3.10: IndianFoodNet dataset Sample of test images dataset.

In addition, it enabled random rotation at a right angle in the clockwise, counterclockwise, and even inverted orientations, improving the images' mobility. The augmenting provides the model with a greater variety of example images, making it feasible. It is, therefore, both reliable and generalizable. The offline augmentation and the enhanced samples were produced simultaneously throughout the training phase. Then, the enhanced samples enhanced the model's replication potential. At the same time, expenses and training time decreased.

3.4.3 Influence of the Dataset

Through analysis, it was discovered that one of the earliest studies using the current DL computer vision technology was IndianFoodNet. Moreover, feeding the model an abnormally high number of food images is unnecessary. It takes less time to train than other models. The problem's emphasis is more precisely defined. Concurrently, this also provides opportunities for further study in computer vision (DL), as food-related visual inputs are highly challenging.

A food-detection model is trained using faster R-CNN training techniques and YOLO variations using the custom-built dataset. Different YOLO variations have been created, and every single one generates a distinct level of efficiency and recognition precision.

3.4.4 Summary

This chapter aims to determine the nutritional value of different food products and train a model to recognize them using a DL algorithm. A custom dataset was created since, in previous work, an image dataset appropriate for the research requirements could not be found. Initially, a list of commonly consumed food items in India from different regions was compiled. After that, a wide range of border areas were utilized to determine the images of meals that were obtained from each food item in each image. After processing, more than 5000 images were tagged with more than 15,000 food items from 30 different cuisine classes, creating the IndianFoodNet dataset. Because of their combined high accuracy and speed, the YOLO5, YOLO7, and YOLO8 algorithms may be trained and compared using this special dataset.

3.5 PHASE 3

Many people like to eat fast food and take in sugar-filled soft drinks, which are heavy in nutrients. In today's world, staying healthy is a top priority for many due to the risks of obesity and related illnesses. Obesity often stems from an imbalanced diet, characterized by high-calorie, low-nutrient foods, and a lack of physical activity. People of each age suffer from this global epidemic, which has a chance to worsen critical illnesses including heart attacks, diabetes, and high cholesterol. Fat occurs primarily from eating a lot of food and not much or at all exercising [104, 105]. Therefore, it is essential to monitor the diet properly [106]. One challenge in managing diet and nutrition is accurately measuring food intake. Many struggle to gauge portion sizes and understand the nutritional content of what they eat. Every person needs a system that tells them about food quantities and calories and gives them instructions on how to eat them. While having a personal nutritionist would be ideal, it's not feasible for everyone. Therefore, there's a need for an automated system that can provide instant guidance on food intake based on images of meals. DL is a powerful image processing method that addresses problems and maintains nutrition since deeper networks can process several features in an image more efficiently.

In this study, a hybrid architecture is developed to assess the calorie content of different food items within a lunch bowl by utilizing deep neural networks. Three essential parts make up this architecture: volume and calorie computation, segmentation, and classification. First, the images of food bowls are segmented using a mask R-CNN to identify individual items. Next, each food item is classified by extracting characteristics from these segmented images using the YOLO V5 framework. Once identified, the dimensions of each item are determined, enabling the calculation of its volume. Lastly, each food item's calorie amount is calculated using its volume. This method can reliably detect and forecast the calorie content of different foods since it has been trained on a library of food images. It aims to provide personalized dietary guidance based on individual food intake. This method's efficacy is assessed and contrasted in conjunction with additional categorization techniques, including CNN, YOLO, and Mask (R-CNN), based on several parameters, including accuracy, sensitivity, and precision.

3.5.1 Existing Algorithms

We have evaluated and contrasted the suggested method's effectiveness with various classification methods, like CNN, YOLO, and Mask (R-CNN). This section elaborates on the concept of the existing algorithm. The detailed explanation of the proposed methodology follows in the subsequent section.

Mask R-CNN technique

Mask R-CNN is one method for identifying and segmenting objects. The objectives of object identification and segmentation are to locate a bounding box around a specific object and identify different objects in an image. Not only does it generate a bounding box around the target item, but by classifying and identifying the pixels inside the bounding box as belonging to it, it can also identify the object, show its border, and highlight significant spots.

Image segmentation is one area where Mask R-CNN, a quicker R-CNN variation, finds application. Figure 3.11 shows the network architecture of the system. Mask (R-CNN) and Faster R-CNN use region proposal network (RPN) for feature collection, limit box classification, and strengthening. Using ROIPool as a feature extraction approach [107], Faster R-CNN leverages max pooling to solve the issue of RoI feature sizes at different scales, allowing any ROI area to be quantified. But in the process, spatial information is lost, which causes the ROI and the original image's extraction attributes to shift. To solve this issue, Mask R-CNN uses ROI alignment (ROIAlign) rather than ROI pooling from Faster R-CNN. Next, the ROIAlign result is displayed over the object region via the mask branch. After the network design was completed, Mask R-CNN was trained using photos. During the instruction phase, the collected instances were split into two stages at random: a training collection and an evaluation set. To verify the model's accuracy and stability, the data from the instruction set was used to test it against the validation set. By lowering the value of the loss factor L in the mask R-CNN, the NN type was found. Based on the training data, the NN model was determined to be the best method. The validation set was one of the fresh data sets the trained model used to forecast and assess.

Mask R-CNN's loss function is described as follows:

$$L = L_{class} + L_{box} + L_{mas} \tag{3.1}$$

When recognized similarly to Faster R-CNN, they are described as:

$$L_{class} + L_{box} = \frac{1}{N_{cls}} \sum_i L_{cls}(\mathrm{P}_i, \mathrm{P}_i^*) + \frac{1}{N_{box}} \sum_i \mathrm{P}_i^* L_1^{smooth}(t_i - t_i^*) \tag{3.2}$$

$$L_{cls}(\{\mathrm{P}_{i,}, \mathrm{P}_i^*\}) = -\mathrm{P}_i^* \log \mathrm{P}_i^* - (1 - \mathrm{P}_i^*) \log(1 - \mathrm{P}_i^*) \tag{3.3}$$

The average binary cross-entropy loss is also calculated as follows:

$$L_{mask} = -\frac{1}{m^2} \sum_{1 \le i, j \le m} [y_{ij} \log oy_{ij}^k + (1 - y_{ij}) \log(1 - o_{ij}^k)] \tag{3.4}$$

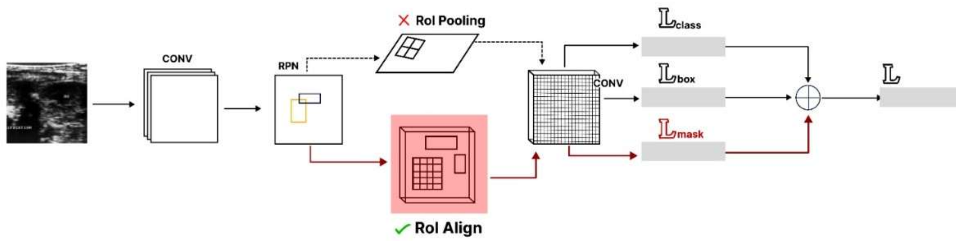


Figure 3.11: The network architecture of Mask R-CNN

Its trained Mask R-CNN model's performance was measured using the accuracy of division on the test set, or mAP:

$$mAP = \frac{A \cap B}{A \cup B} = \frac{1}{N^T} \sum_{i=1}^{N_T} \left(\frac{N_i^{DR}}{N_i^D} \right) \tag{3.5}$$

Accuracy served as a validation measure for the method's overall classification performance. The following formulas are used to estimate the accuracy measure:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{3.6}$$

where FN = false negative, TN = true negative, TP = true positive, and FP = false positive.

Algorithm 3 Mask R-CNN Object Detection and Segmentation

Input: Image dataset $D = \{I_1, I_2, \ldots, I_n\}$.
Output: Segmented objects with bounding boxes and masks.
Step 1: Feature Extraction
Use Region Proposal Network (RPN) to generate region proposals $R = \{r_1, r_2, \ldots, r_k\}$ from image I.
Extract features from I using ROIAlign for each proposal r_i.
Step 2: Object Detection
Classify object in each proposal region r_i and predict bounding box using:

$$L_{\text{cls}}(p_i, u) = -\log p_{u,i} \tag{3.2}$$

Predict bounding box offset using regression:

$$L_{\text{box}}(t_i, v) = \sum_{i \in \{x,y,w,h\}} \text{smooth}_{L_1}(t_i - v_i) \tag{3.3}$$

Step 3: Mask Prediction
Predict object mask over region r_i using binary cross-entropy loss:

$$L_{\text{mask}} = -\frac{1}{m^2} \sum_{i,j} \left[y_{ij} \log(p_{ij}) + (1 - y_{ij}) \log(1 - p_{ij}) \right] \tag{3.4}$$

Step 4: Model Training
Split data D into training set D_{train} and validation set D_{val}.
Train Mask R-CNN on D_{train} and validate on D_{val} by minimizing the total loss:

$$L = L_{\text{cls}} + L_{\text{box}} + L_{\text{mask}} \tag{3.1}$$

Step 5: Model Evaluation
Evaluate performance on test set using mean Average Precision (mAP):

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \tag{3.5}$$

Calculate accuracy using:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.6}$$

End.

Algorithm 3 describes step by step how to carry out object detection and segmentation using Mask R-CNN on an image set. It includes region proposal of an image through a Region Proposal Network (RPN), proposing regions for classifying objects and estimating bounds, generating predictions about the object's masks, training a Mask R-CNN on a training data set, and assessing the performance through mean Average Precision (mAP) and accuracy ratio.

YOLO

First, the source image is split into a pattern of small individual cells to begin the YOLO detection procedure. It is the job of each cell to determine whether or not it contains an item and, if so, what kind of object it may be involves predicting bounding boxes (like boxes around objects) and the probability of each box containing a certain type of object. Each prediction comes with a confidence score, indicating how sure the model is about its prediction. To avoid duplicate or unnecessary predictions, non-maximum suppression is used. The final result shows the detected objects, their types, and how confident the model is about its findings. This process is repeated for each frame in a video, allowing for real-time object detection. The candidates box extraction, feature extraction, and object categorization techniques are all included in the YOLO neural network. Candidate boxes are immediately extracted from images using the YOLO neural network, and objects are identified by utilizing all of the image's attributes. Figure 5.2 depicts the YOLO detection procedure. The supplied image is first scaled. After that, an image is a single convolutional network, yielding border boxes and confidence ratings. Finally, the result class rates and border boxes are found in the top right of Figure 3.12, using an NMS to cut the ensuing results by the model trust.

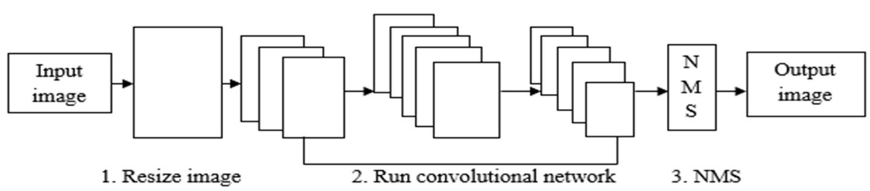


Figure 3.12: The detection process of YOLO

YOLO is used to divide the image given into 7x7 cells. Every matrix pixel has two bounding boxes corresponding to it. Every grid cell forecasts a single set of category chances, including 20 classes, regardless of the number of boxes. A limit box also contains five predictions: x, y, w, h, and trust. The (x, y) values indicate the base of the frame that is closest to its boundaries; its width and height in relation to the whole image are given by the (w, h) values. Thus, there are 30 predictions for each grid $(4 + 1) \times 2 + 20 = 1470$ predictions for the entire image $(7 \times 7 \times 30 = 1470)$.

Yolo's CNN consists of 25 layers of convolution, 4 merging levels, 1 dump level, 1 completely linked level, and 1 identification level. The completely connected layer has a size of 1470, which represents the class probabilities, anticipated bounding box position, and confidence. The output of the entire layer of connections is also random when a CNN is first begun, which means that there is initially a major gap between the forecast container and the actual reality box. The ground truth box must be approached over several iterations, which slows the velocity of closure. Nevertheless, YOLO's localization error is substantial as it proposes far fewer bounding boxes only $7 \times 7 \times 2 = 98$/per image.

CNN

CNN is a DL technique used for object identification, feature extraction, semantic segmentation, and image classification [108]. Due to its large network depth and widespread application with image data, CNN is classified as a deep neural network [109]. CNN is applied to object detection and grouping of images [110]. CNN is able to recognize objects because it employs the method of convolution to provide the computer with information. A technique used to recognize and classify various food products in images or videos is called object detection for food object classification. CNN's two main components are the fully linked layer and the extraction layer. Feature extraction consists of a convolutional layer and a pooling layer.

Convolution Layer

By doing the convolution process with the mask and the picture, the mask layer helps to determine the texture or pattern in the image. In order to analyse a 5x5 image, as illustrated in Figure 3.13, a 1-pixel step or twist lens is applied within the original image and the 3x33 lens is placed in the layer of convolution, which is situated on the primary layer of the Dreams architecture.

According to the YOLO algorithm in Algorithm 4, this paper describes how object identification occurs in images and videos. It implies splitting the image into a grid, estimating the coordinates of the boxes and likelihoods of each cell, applying suppression to filter many similar instances, and rendering the detected objects with their orientation boxes and confidence level. In this case, the classifiers used incorporated a convolutional neural network structure, which has a fully connected layer used to predict these values. In the case of video object detection, the same process is performed frame by frame on the sequence of video frames.

Algorithm 4 YOLO Object Detection Process

Input: Image I.
Output: Detected objects with bounding boxes and confidence scores.
Step 1: Image Division
Divide image I into 7×7 grid cells.
Each grid cell predicts:

$$B = \{(x, y, w, h, c)\} \quad \text{for two bounding boxes}$$

where x, y are coordinates, w, h are width and height, and c is confidence.
Step 2: Class and Bounding Box Prediction
For each grid cell, predict class probabilities for 20 classes:

$$P(C) = \{p_1, p_2, \ldots, p_{20}\}$$

Each grid cell outputs 30 predictions:

$$30 = (4 + 1) \times 2 + 20$$

This results in 1470 predictions for the entire image ($7 \times 7 \times 30$).
Step 3: Non-Maximum Suppression (NMS)
Apply NMS to reduce overlapping and redundant bounding boxes based on confidence scores c.
Step 4: Final Output
Output bounding boxes, class probabilities, and confidence scores for the detected objects.
Step 5: Video Object Detection
Repeat the above steps for each frame of a video sequence for real-time detection.
Step 6: Neural Network Architecture
YOLO's CNN consists of:

$$\text{25 convolutional layers}, \quad \text{4 merging layers}, \quad \text{1 drop layer}, \quad \text{1 fully connected layer}$$

The fully connected layer has 1470 neurons to predict bounding box positions, class probabilities, and confidence scores.
Step 7: Ground Truth Optimization
Iteratively adjust bounding box predictions to match the ground truth over several epochs to minimize localization error.
End.

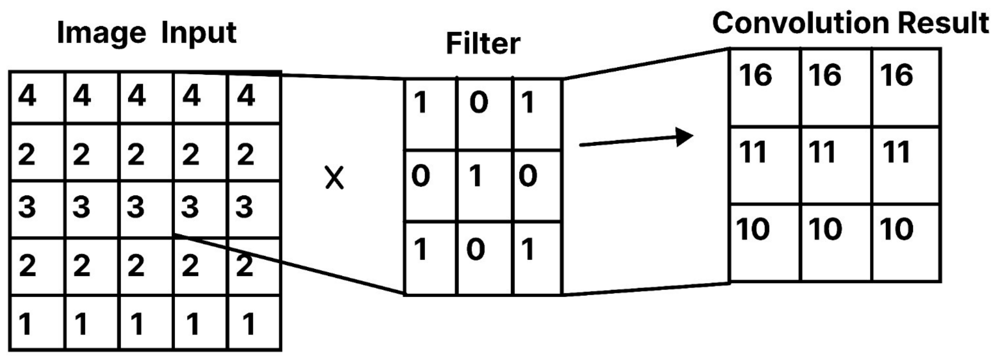


Figure 3.13: Convolution layer process

The input image is divided by the mask. The output of the convolution layer is then displayed on the right side. The results obtained by filter multiplication of the input images will be handled by the layer for pooling.

Pooling Layer

It is the layer for pooling that partitions or divides many pieces into a square form as displayed in Figure 3.14. The layer will after that use the finest value of each in the inversion layer's output for pooling. Using the max-pooling filter $2x2$ and step 1, this layer will shift the mask by 1 pixel in order to find the maximum value.

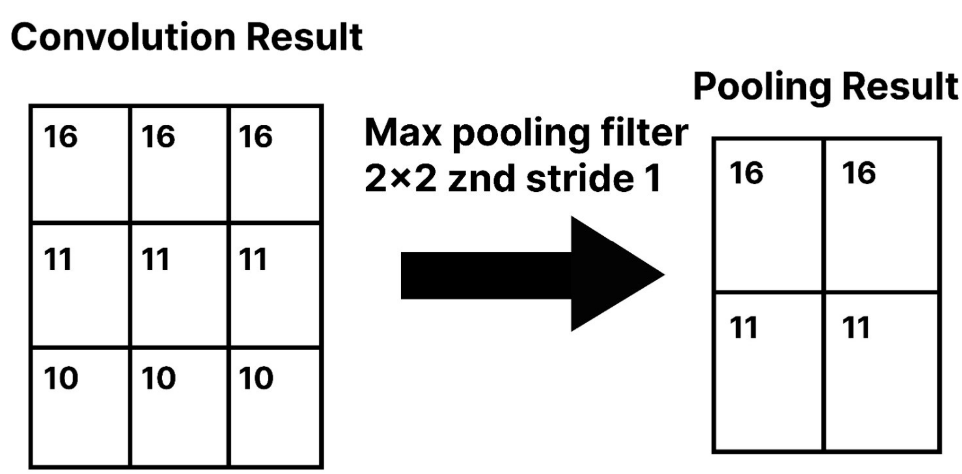


Figure 3.14: Pooling Layer Process

Fully Connected Layer

The fully connected layer is the final layer. This layer connects other neurons to a single neuron and stores the form of that neuron in order to collect the findings of the whole layer. When the outcomes of the preceding layer are arranged in the completely linked layer. Acquired are the pooling computation results. These results are then treated as two vectors, denoted as x and y. Next, during the softmax stage, the probability obtained from the individual vectors is evaluated against the category whose services the model has generated. Equation 3.7 may be used to see the computation as follows:

$$\begin{aligned} \sigma(z_1) &= \frac{e^x}{e^x + e^y} \\ \sigma(z_1) &= \frac{e^y}{e^x + e^y} \end{aligned} \tag{3.7}$$

The vector is where the values are found. Because the food dataset has heterogeneity in appropriate image processing, images are shot at the ideal image for evaluation, maintaining identical brightness and range.

3.6 SUMMARY

The study suggested a hybrid approach for evaluating food products' calorie content using DL algorithms. First, segmentation was applied to the images using Mask (R-CNN). After segmentation, food items were categorized using the YOLO V5 framework after characteristics were retrieved from the segmented images. Following the identification of the food items, the dimensions of each were ascertained, and these estimates were used to determine the quantity of each kind of food. The volume of the food items was then used to determine the calorie amount. The models stated above were created using the dataset's food images.

CHAPTER 4

EXPERIMENTAL SETUP AND FRAMEWORK

4.1 INTRODUCTION

The previous chapter deals with different phases of research and describe proposed methodology. This study suggested a hybrid approach for evaluating food products' calorie content using Deep Learning algorithms and implementation of calorie estimation.

4.2 IMPLEMENTATION OF PROPOSED METHODOLOGY

An implementation of proposed methodology is displayed in Figure 4.1. The present research examines the suggestion that was made using the Indian Food Net-30 dataset. Particular images of the food in the dish are part of the dataset. Subsequently, the food item images are segmented using the mask (R-CNN). The meal products are classified after features are gathered from the divided images using the YOLO system. Each food product's dimensions are established after they have been identified. The estimated dimension is used to compute the food item's volume. Next, the food product volume serves to estimate the calorie content.

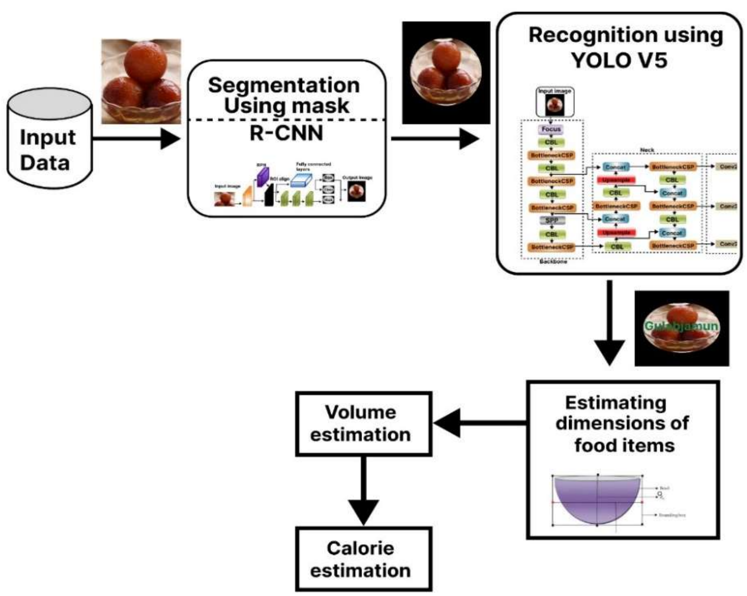


Figure 4.1: The overview of the proposed scheme

Calories are being calculated from food images by proposed scheme. The psuedocode to find total calories are formulated as follows:

1. Food Images(F)= (F_1, F_2, ……..F_n) ∈ $Class_i$, i⟵30
2. Labelled Images (Ln)
3. Ln is split into Training and Testing set: $Iset = Trainset \cup Testset$
4. MaskRCNN applied to $Trainset$ for Instance Segmentation:

   $Output_1 = BBo(ROI) \wedge Mask(ROI)$
5. Classification done by pre-trained network – YOLOv5:

   $Output_2 : Ln = Class_i$
6. ROI Mark Dimensions: $d_a = Height_i$ and $d_b = Diameter_i$
7. Estimation of Volume: $Volume_i = \pi /6\ (3(dia_i)^2 + (h_i)^2)\ h_i$
8. Estimation of Calorie: $Total_{Calorie} = Vol_i \times C_E$

   The pipelining of deep learning for the evaluation of the overall calorie content from food images are described in Algorithm 5. It includes naming the food images, partitioning the dataset into train and test sets, instance segmentation using mask R-CNN to obtain the regions of interest, food item classification using YOLOv5 for obtaining the coordinates of the food items, measuring object dimensions, estimating the volume of each item and finally the total calorie estimate of the meal.

4.3 DATASET DESCRIPTION

“Ritu Agarwal, Tanupriya Choudhury, Nikunj Bansal, Tanmay Sarkar and Neelu Jyothi Ahuja” developed IndianFoodNet-30, the dataset of over 5,500 images of 30 common Indian meals, to create an Indian food identification algorithm. “Chole, Bhindi Masala, Ras Malai, Samosa,Biryani, Chai,Gulab Jamun, Idli, Jalebi, Aloo Gobi, Aloo Masala, Bhatura, Lassi, Mutton Curry, Onion Pakoda, Palak Paneer, Poha, Rajma Curry, Shahi Paneer, Kebab, Kheer, Kulfi,Coconut Chutney, Dal, Green Chutney, White Rice, Dosa, Dum Aloo, Fish Curry, Ghevar”, Furthermore, all of the collection's images have been resized to 640 by 640. The model in this study is trained using eighty percent of the data. The trained model is tested using the remaining components of the model.

Algorithm 5 Calorie Measurement Using Deep Learning

Input: Food images $F = \{F_1, F_2, \ldots, F_n\}$, where $F_i \in \text{Class}_i$, $i \in [1, 30]$
Output: Estimated total calorie content
Step 1: Image Labeling
Obtain labeled images L_n.
Step 2: Dataset Splitting
Split L_n into training set and testing set:

$$\text{I}_{\text{set}} = \text{Train}_{\text{set}} \cup \text{Test}_{\text{set}}$$

Step 3: Instance Segmentation
Apply Mask R-CNN to $\text{Train}_{\text{set}}$:

$$\text{Output}_1 = \text{BB}(ROI) \wedge \text{Mask}(ROI)$$

Step 4: Object Classification
Use pre-trained YOLOv5 to classify food items:

$$\text{Output}_2 : L_n = \text{Class}_i$$

Step 5: Dimension Extraction
Extract object dimensions:

$$d_a = \text{Height}_i, \quad d_b = \text{Diameter}_i$$

Step 6: Volume Estimation
Estimate the volume of food item i:

$$\text{Volume}_i = \frac{\pi}{6}\left(3 \cdot (d_b)^2 + (d_a)^2\right) \cdot d_a$$

Step 7: Calorie Estimation
Estimate the total calories based on volume:

$$\text{TotalCalorie} = \sum_{i=1}^{n} \text{Volume}_i \times CE$$

where CE is the calorie estimate per unit volume.
End.

4.4 CALIBRATION

Calibration is crucial in image processing to ensure accurate measurements, particularly when dealing with real-world dimensions [111]. Several widely used techniques of calibration in image processing and computer vision include:

a. Camera Calibration [88]

Finding the extrinsic and intrinsic properties of the camera is necessary for camera calibration. This assists in converting image coordinates to real-world coordinates and correcting lens distortion.

Steps for Camera Calibration

- Use a known pattern, such as a checkerboard, coin, or grid. Capture multiple images of this pattern from different angles and positions [112].
- Use functions like cv2.findChessboardCorners to detect the corners of the pattern in each image.
- To calculate the camera matrix, rotation, translation vectors, and distortion coefficients, use cv2.calibrateCamera.

b. Scale Calibration

Scale calibration ensures that the measurements in the image correspond to real-world units. It involves using a reference object of known size within the image [113].

Steps for Camera Calibration

- Include an object with known dimensions in the scene.
- Measure the pixel length of the reference object in the image.
- Compute the scale factor to convert pixel measurements to real-world units.

The process of estimating a camera model's parameters, or camera calibration, enables the correction of lens distortions, the calculation of the camera's orientation, and the comprehension of the correspondence between 2D image coordinates and 3D real-world coordinates.

The key steps and concepts involved in camera calibration include:

1. **Intrinsic Parameters**: These parameters specify the focal length, optical center (principal point), and lens distortion coefficients, among other internal camera properties.

- **The Focal Length (f_x, f_y)**: The distance between the lens and camera sensor.

- **The Principal Point (c_x, c_y)**: It is a point where the optical axis intersects the image plane.
- **Distortion Coefficients**: Parameters that describe the lens distortion, including radial and tangential distortions.

1. **Extrinsic Parameters**: These parameters define the orientation and camera's position in the world of coordinate system.

- **Rotation Matrix (R)**: Describes the camera orientation.
- **Translation Vector (T)**: Describes the camera position.

2. **Calibration Process**:

- **Image Acquisition**: Take several pictures of a known calibration pattern—such as a chessboard—from various viewpoints and locations.
- **Feature Detection**: Detect the known points (e.g., corners of the chessboard) in the images.
- **Parameter Estimation**: Utilize procedures (such as Zhang's method) to minimize the reprojection error—the discrepancy between the projected 3D coordinates and the observed image points—in order to estimate the intrinsic and extrinsic parameters.
- **Extracting 3D Information from 2D Images [89]**

Once the camera is calibrated, it is possible to extract 3D information from 2D images. The main techniques include:

1. **Stereo Vision**: Using two cameras placed at different positions to capture the same scene from different viewpoints. By matching points in the two images (stereo matching), the depth (distance from the cameras) can be calculated using triangulation.
2. **Structure from Motion (SfM)**: Using multiple images taken from different viewpoints, the 3D structure of a scene can be reconstructed. This involves:

- **Feature Detection and Matching**: Detecting and matching key points across the images.
- **Camera Pose Estimation**: Estimating the relative positions and orientations of the camera for each image.
- **3D Reconstruction**: Reconstructing the 3D points by triangulating the matched features.

3. **Depth Sensors**: Devices like LiDAR or depth cameras (e.g., Microsoft Kinect) provide direct measurements of depth information for each pixel.

4.5 SEGMENTATION USING MASK (R-CNN)

Following dataset construction, pictures are sent to Mask (R-CNN) for categorization. A DL-driven method for instance classification and object identification is the Mask R-CNN. Different masks are assigned to distinct instances of the same class of objects through the use of instance segmentation. There are two separate parts to the structure of the framework. In the first stage, the Region Proposal Network (RPN) pre-trained network uses ResNet to generate a probability for the positions of the food items in the image.

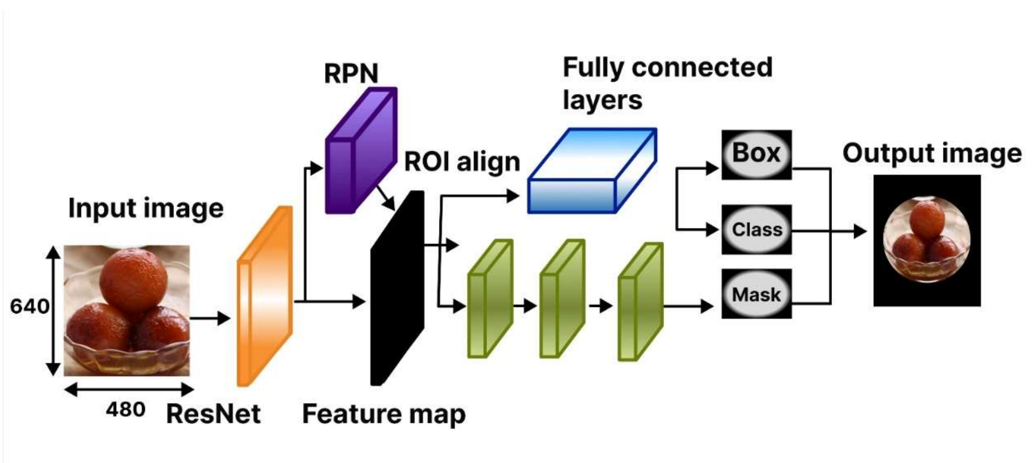


Figure 4.2: Architecture of Mask R-CNN

The Faster R-CNN object identification technology, which consists of the RPN and the Fast R-CNN network, provides the foundation for the Mask R-CNN architecture [114].

Mask R-CNN further contains a branch called division in addition to these elements. Figure 4.2 depicts the Mask R-CNN design.

According to the present Faster R-CNN architecture, the third branch is at its top and uses the identical getting features core as the previous two divisions. A portion, called the RPN, slides an isolated network created by the main one across a convolutional map to provide area concepts.

These region recommendations are fed into the second component, the Fast R-CNN network, which performs limit box regression and classification to produce item detections. For instance, the third branch generates binary masks indicating that each offer's pixel values correspond to the item of interest by using the same region proposals as input. A fully convolutional network (FCN) that integrates values from the previous two versions is used to construct this branch [115]. The following is an explanation of the parts of MASK (R-CNN).

4.5.1 ResNet-FPN

Mask-R-CNN uses ResNet-FPN as its foundational architecture for extracting features. A deep CNN called ResNet has achieved remarkable performance in picture classification. ResNet-FPN builds a more powerful feature extractor by fusing the ResNet base with FPN architecture. Object recognition accuracy can be increased by using the FPN component to create multi-scale feature maps from an input image.

4.5.2 RPN

The features retrieved with ResNet101+FPN are sent into the RPN network as input to generate ROIs. If proportions of crease size to width alter, RPN may be able to distinguish between an image's front and back. In this instance, insert the image box on the connection and give the characteristic you want in the image's surround box specifications to rapidly produce candidate areas. The related pegs are created by a 3x3-deep layer that scans the images and places the stakes in multiple sizes near the image.

A web page adjusts its size according to the images provided and has several anchorages at predetermined points. Every limit box and ground-truth category have an anchor attached to it. Several basic limit shapes and aspect ratios are available to designate flaws of different shapes and sizes. The fact that these boxes overlap several times facilitates the identification of the maximum confidence score for multiple ROI detection. An equation called the intersection-over-union factor (IoU) calculates it.

$$IoU = \frac{Areaof\ Overlap}{Areaof\ union} \tag{4.1}$$

4.5.3 RoIAlign

Rectangular region suggestions are sent into RoIAlign, which then pulls features regarding the element of the map a given idea is associated with. For an R-CNN network to enable mask branch recognition of pixel targets, it must have superior visual accuracy and the ability to determine if an extension belongs to a distinct visual goal. The original image is pooled and convolutioned, then its size is altered. The next step is segmentation. Thus, this work proposes Mask (R-CNN), an enhanced form of Faster (R-CNN). Furthermore, RoIAlign is used as CNN's pooling layer. This method uses linear interpolation to maintain the spatial intricacies of the feature map. A neural network layer called RoIAlign is utilized in instance segmentation methods such as Mask R-CNN and object identification.

4.5.4 FCN

FCN controls the division of the filter applied to the region of interest (ROI). FCN makes predictions pixel-by-pixel. The FCN model has implemented the per-pixel soft maximum loss calculation.

Loss Function

The Mask R-CNN technique specifies the multitask loss function () for each RoI gathered throughout instruction, which is given as

$$Lf = Lf_{cls} + Lf_{box} + Lf_{mask} \tag{4.2}$$

The symbols for segmentation error, detection error, and classification error are as follows:

Within Mask (R-CNN), they are shown as

$$Lf' = \frac{1}{W_{cls}} \sum_i Lf_{cls}(p_i, p_i^*) + \lambda \frac{1}{W_{reg}} \sum_i p_i^* Lf_{reg}(v_i, v_i^*) \tag{4.3}$$

where indicates the estimated likelihood that the $_i$-th target will be on the anchor point. It is determined using the anchor point sample's sign. is 1 or 0, depending on whether There is a favourable test located at the pin location.

The vectors, which represent the amount of change in the positive sample anchor point with respect to mark off the zone and the expected area, respectively, are composed of four translational and scaling components. Weight and manage the two losses in order to keep everything in balance. The following are the regression losses and classification losses:

$$Lf_{reg}(v_i, v_i^*) = smooth_{Lf}(x)(v_i, v_i^*) \tag{4.4}$$

$$Lf_{cls}(p_i, p_i^*) = -\log\left[p_i, p_i^* + (1 - p_i^*)(1 - p_i)\right] \tag{4.5}$$

The x translation of the rectified frame at the horizontal axis anchor point defines the robust loss. It's characterized as

$$smooth_{Lf}(x) = \begin{cases} 0.5x^2, & |x| < 1, \\ |x| - 0.5, & otherwise \end{cases} \tag{4.6}$$

Mask R-CNN, a typical dual-entropy operation, explains the reduction of semantic division splits [115]. After handling, the feature map provided in a mask arm is put in a file where the area and size of its features are set by or, respectively.

Computing the pixel-by-pixel sigmoid of the resultant feature map yields the average entropy error and relative entropy. Consequently, the proposed mask (R-CNN) successfully divides the meal products' locations within the image into parts.

4.6 IDENTIFYING FOODS WITH YOLO V5

The YOLO V5 predictor receives the divided images after that for additional processing categorization after segmentation. After characteristics are collected from the segmented photos, the food items are identified using the YOLO V5 network. The first single-stage recognition methodology, the Yolo item identification technique, was created by Redmon J. After all of the images have been fed to the network, the basic concept of YOLO, or Faster (R-CNN), is to immediately reverse the types and locations of the bounding boxes on the output layer. Regression problems that incorporate bounding box and classification problems do not require the candidate box extraction stage of the two-stage approach. Version 5 of the YOLO network keeps the advantage of fast detection speed while achieving far higher detection accuracy than previous versions. This lessens the negative effects of the low precision of detection of the one-stage system. Figure 4.3 displays the YOLO V5 network architecture. There are three main parts to the YOLO:

Backbone: CNN uses image resolution to compile and generate image attributes.

Neck: A group of computing levels that, prior being forwarded to the forecast layer, integrate and aggregate image attributes.

Head: It can anticipate image properties and generate border boxes and classifications. The level of confidence shows how correctly a classification was formed in a particular situation.

The three parts of the yoloV5 structure are the neck, output, and backbone. The input image at the appropriate resolution is received by the backbone's focus component. A basic convolution module is the CBL. A CBL module is represented by the symbols Batch Normal, Leaky RELU, and Conv2D. In order to extract rich data from the image, the Bottleneck CSP module efficiently executes feature extraction to the feature map. For the entire unit, the variable's value is mostly determined by its parameter amount.

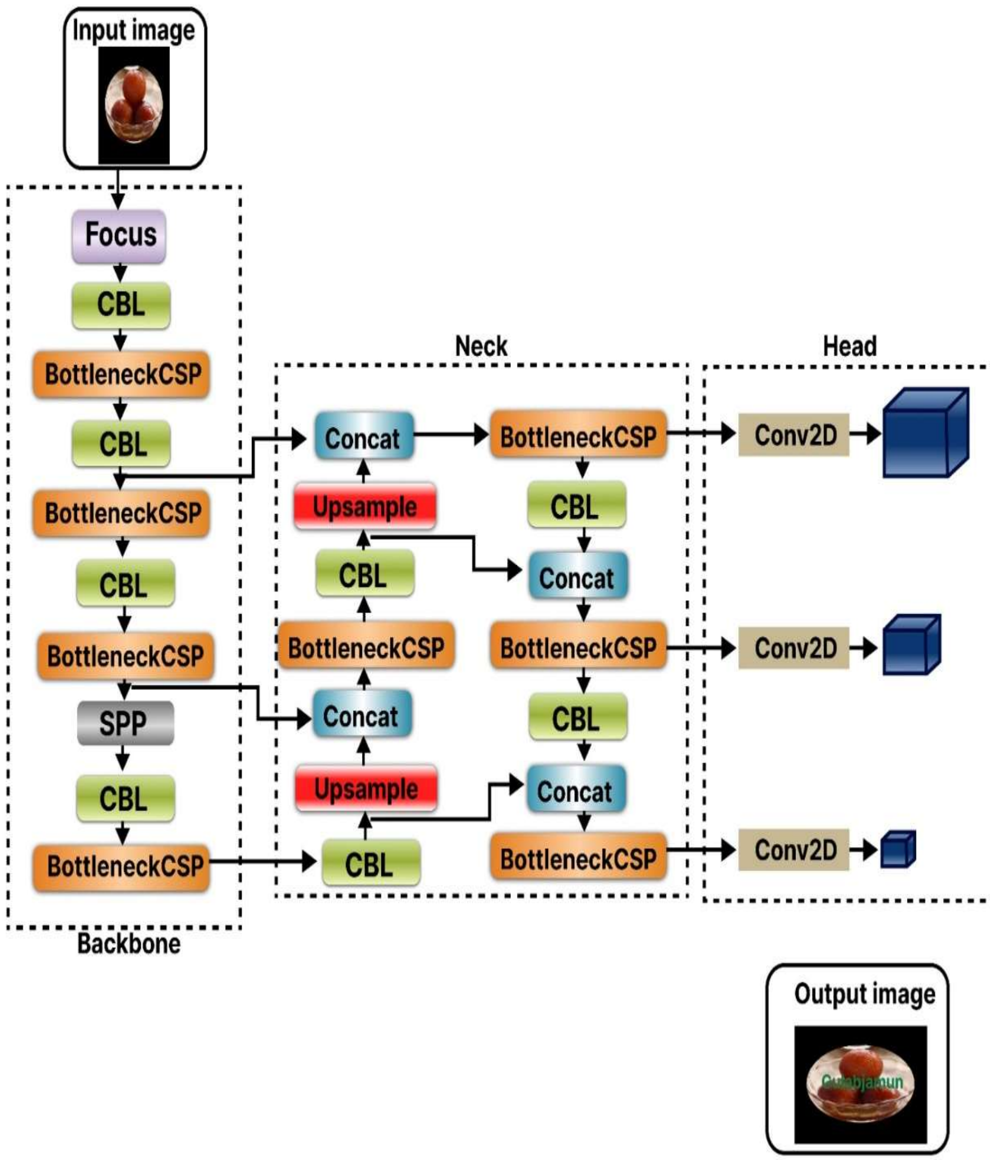


Figure 4.3: YOLO V5 network structure

The SPP module serves multiple functions and increases the network's receptive area. Additionally, YOLOv5 has an aspect of structure that is bottom-up and built on the FPN architecture. The FPN layer transfers powerful conceptual data compared to the highest point to the bottom, whereas the feature triangle transmits firm place features from the bottom up. At the base of the graphic, show the item coordinates and the classification outcomes.

Input

Three components comprise the input: data enhancement, image dimensions analysis, and a holding structure that dynamically adapts. The classic YOLO v5 uses patchwork data augmentation to arbitrarily size, trim, set up, and then fuse the input to increase the detection of tiny objects [116]. During dataset training, the size of the provided image is changed to a standard size before being fed into the model for analysis.

Backbone

The network is built on the focal point and CSP architecture. The Focus structure slices the image before it enters the backbone network. The original picture was split, as Figure 4.4 illustrates. The feature map is constructed using the convolutional algorithm once it has reached half of the image's solution. Without requiring any adjustments to the parameters, the Focus can reduce the input sizes while maintaining as much of the original image information as is practically possible. [117]. The CSP architecture is created from the input feature using two 1*1 convolutions. It facilitates the removal of computational bottlenecks, reduces memory expenses, and improves CNN learning.

Neck

The neck area, an internet layer, provides the collected image information to the estimation layer. The neck section of the YOLO V5 uses the FPN+PAN. To generate the feature map required for forecasting, the FPN integrates top-to-bottom great qualities and leverages communication [118]. Strong positional properties are communicated from bottom to top by the underlying pyramid, or PAN.

Output terminal

This estimation layer builds the border box to anticipate the image's attributes and classification. YOLO V5 utilizes CIoU Loss as its loss function for the bounding box. Compared to traditional non-maximum suppression (NMS), CIoU_NMS is more effective at identifying overlapping items.

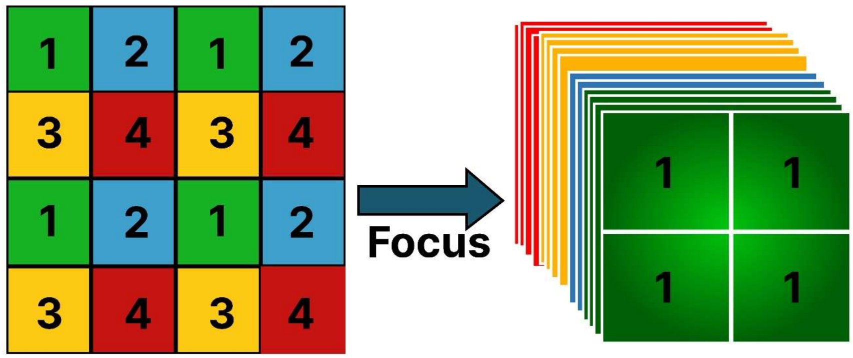


Figure 4.4: The Focus Model for Flow of process.

YOLO v5 inherits the connection concepts from the original YOLO algorithm. The original image size was processed as a 3-channel RGB colour image. The output dimension as the three enormous, medium, and micro scales enters the systemin detecting layer is,

$$S \times S \times n_a \times (t_x + t_y + t_w + t_h + t_o + n_c) \tag{4.7}$$

$S \times S$ is the divided mesh's number. The sign on each scale represents the number of preset preceding boxes. There are several tasks that call for forecasting. With a 9600output category from the recognizing layer, the large-scale are examples of network structures.

Bounding box parameters are related as follows: The bounding box's confidence level is. After that, category I's confidence is. To obtain the final estimation box, these elements must be decoded using the following formulas:

$$a_x = 2\sigma(t_x) - 0.5 + c_x \tag{4.8}$$

$$a_y = 2\sigma(t_y) - 0.5 + c_y \tag{4.9}$$

$$a_w = p_w (2\sigma(t_w))^2 \tag{4.10}$$

$$a_h = p_h(2\sigma(t_h))^2 \tag{4.11}$$

$$score_i = confidence \times \Pr(class_i) = \sigma(t_o)\sigma(t_{ci}) \tag{4.12}$$

The fact box's basis is shown by the blue box in Figure 4.5. The label that encircles the box's dimensions, supervisors, and centre point is given as, and then. The measurements are taken from the grid's leftmost point to the label-enclosing box's centre. The anchor box is the red box. The width and height of the preceding frame are, accordingly.

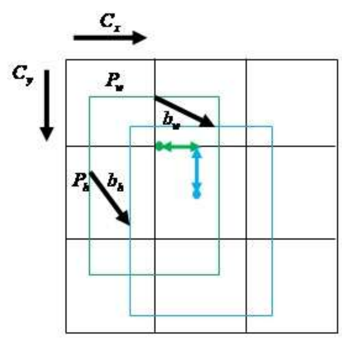


Figure 4.5: Interpretation of the YOLOv5 prediction boundary box

A common issue that significantly impairs algorithm performance while training on a visual identification task is a mismatch between positive-negative (foreground-background) trials. Following development, YOLO v5 allocates the label box to three anchors simultaneously, double the desired sample size. During the target identification algorithm's training phase, the difference between each of the positive and negative samples is slightly changed. The loss function is calculated by formula (13). The bounding box's loss ratio is reduced.

$$\begin{aligned} L_{total} &= \sum_{i}^{N}\left(\lambda_1 L_{box} + \lambda_2 L_{obj} + \lambda_3 L_{cls}\right) \\ &= \sum_{i}^{N}\left(\lambda_1 \sum_{j}^{B_i} L_{CIoU_j} + \lambda_2 \sum_{j}^{S_i \times S_i} l_{objj} + \lambda_3 \sum_{j}^{B_i} l_{cls_j}\right) \end{aligned} \tag{4.13}$$

N levels of detection are present. The desired number, B, is represented by the labels on the box in front of it. Numerous meshes make up this scale. is the bounding box that is used in the regression loss computation for each target. For each mesh, the loss of the target object is utilized. The classified loss, which was used to select each target, The weights assigned three losses

4.7 VOLUME ESTIMATION

After the object has been categorized and the ROI has been adequately defined, the item's volume will be measured. The amorphous food item takes on the shape of a bowl, so its volume and capacity are the same. We need to know the measurements of the object and the reference object to determine the bowl's volume. In this case, a coin serves as the sign. For volumetric estimation, boundaries are drawn on the item and the currency. Figure 4.6 displays the reference bowl as well as its height (d_a) and diameter (d_b) measurements.

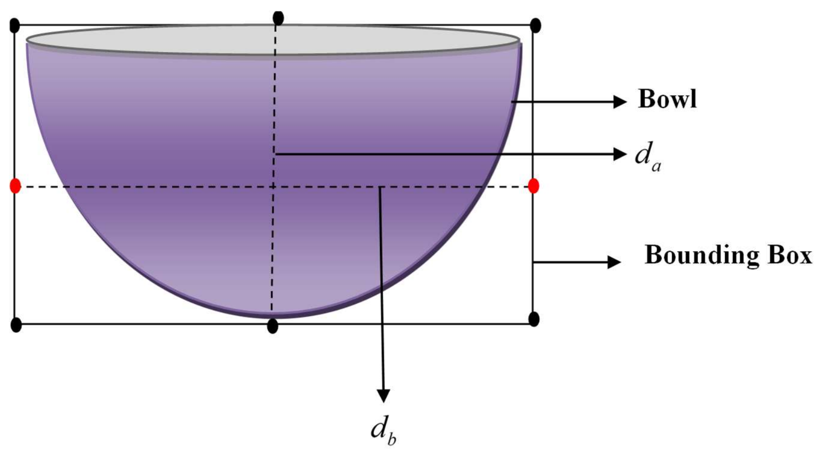


Figure 4.6: Marking the dimensions of the bowl using the pixel approach

The pixel per meter method [119,120] is used to ascertain the quantity of atypical foodstuffs. This method calculates the width in pixels of the reference object. The object's height and width are then calculated using this ratio. Using these measurements, we next find the bowl's estimated volume (Vol $_E$) in Eq. (4.16), which is simply the number of food items shaped like bowls. Equation (4.16) is derived from Equation (4.14), the conventional equation for the hemispherical bowl.

$$Volume_{Bowl} = \frac{\pi}{6}\left(3(r_{Bowl})^2 + (h_{Bowl})^2\right)h_{Bowl} \tag{4.14}$$

$$h_{Bowl} = \frac{d_a}{pPM}$$
$$r_{Bowl} = \frac{1}{2}\frac{d_b}{pPM} \tag{4.15}$$

$$Vol_E = \frac{\pi}{6}\frac{d_a}{pPM}\left(\frac{3}{4}\left(\frac{d_b}{pPM}\right)^2\left(\frac{d_a}{pPM}\right)^2\right) \tag{4.16}$$

$$pPM = \frac{Object-width}{Know-width} \tag{4.17}$$

Here, denotes the width of the bowl, the pixels-per-metric ratio, and the size of the standard item (coin), which is equivalent to 0.9 meters. The symbols (d_a) and (d_b) stand for the Euclidean distance between two points on opposite sides of the objects.

4.8 Calorie estimation

Calorie calculations can be made once the volume is established. Conventional calorie mapping charts are provided for both cooked and raw foods [121]. The graph displays the number of calories per gram. It is converted to grams and linked to the calorie table, which displays the precise number of calories that our system has found a food item to have to establish its exact amount. Calculating total calories and calorie equivalents ($Calorie_{total}$) involves applying the formula below.

$$Calorie_{total} = Volume * C_E \tag{4.18}$$

Therefore, the meal products quantity is used to estimate the energy.

4.9 SUMMARY

The study suggested a hybrid approach for evaluating food products' calorie content using DL algorithms. First, segmentation was applied to the images using Mask (R-CNN). After segmentation, food items were categorized using the YOLO V5 framework after characteristics were retrieved from the segmented images. Following the identification of the food items, the dimensions of each were ascertained, and these estimates were used to determine the quantity of each kind of food. The number of calories was then calculated based on the food's volume. The food photos in the dataset were used to generate the models above. This integrated methodology effectively merges various deep learning techniques to provide a thorough assessment of food calorie content. Below is a brief overview of the process:

Segmentation: Utilizing Mask R-CNN to isolate food items and eliminate background distractions.

Classification: Employing YOLO V5 to identify each food item, thereby determining the type of food present in the image.

Dimension Estimation: Measuring the physical dimensions to approximate the volume of each food item.

Calorie Calculation: Estimating calorie content by correlating food volume with established calorie densities for each food category.

This integrated methodology presents numerous benefits:

Enhanced Accuracy: The combination of segmentation and classification ensures precise identification and quantification of each food item, thereby enhancing the accuracy of calorie estimations.

Robustness: The implementation of both Mask R-CNN and YOLO V5, which are sophisticated deep learning models, allows the system to effectively manage a diverse range of food items and varying environmental conditions.

Scalability: This approach can be expanded to recognize additional food types by training the YOLO V5 model on a larger and more varied dataset. It is also suitable for real-time applications, such as dietary tracking applications.

This integrated methodology holds considerable promise for applications in dietary monitoring, nutritional analysis, and health management, rendering it a valuable resource for both individual and professional health contexts.

CHAPTER 5
RESULTS AND DISCUSSION

5.1 INTRODUCTION

This chapter examines the study's conclusions and arguments. The primary goals were to create a system that efficiently identifies food and calculates its caloric content, as done in the previous chapter. It covers the system's simulation parameters and the assessment metrics that are used to examine its operation. Moreover, thorough explanations of the experimental outcomes obtained in this study are presented, along with an analysis of the model's performance. The proposed methodology is proved to be comparatively efficient for calorie estimation based on the statistical analysis performed using t-test.

5.2 SIMULATION SETTINGS

The simulation parameters for the proposed technique are configured as follows. Python is used to accomplish the suggested method. The system requirements for the implementation are 4GB of RAM and a 1.6 GHz Intel Core i5 system in useWindows. The Roboflow model's GoogleColab Notebook (https://universe.roboflow.com /indianfoodnet/indianfoodnet) is used for testing and training. Included are notebooks related to YOLO variations that were taken from the original source. Table 5.1 shows the settings that were applied throughout the experiment.

Table 5.1: Parameters for Simulation

S.No	Simulations	Parameter
1	Dataset For Testing	20%
2	Epochs	25
3	Device	CPU
4	Optimizer	Adam
5	Split_Ratio	0.2
6	Weight_Decay	1e-3
7	Dataset For Training	80%
8	Batch_Size	2
9	Learning Rate	0.001

5.3 EVALUATION METRICS

Object detection models are evaluated using a variety of performance measures, including recall, accuracy, specificity, precision, MCC, NPV, and error metrics. This inference rate also determines the non-maximal attenuation (NMS). These are the equations for metric assessment mentioned above:

$$\Pr ecison = \frac{TruePositive}{TruePositive + FalsePositive} \tag{5.1}$$

$$\text{Re}call = \frac{TruePositive}{TruePositive + FalseNegative} \tag{5.2}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_{i} \tag{5.3}$$

5.4 EXPERIMENTAL RESULTS FOR PHASE 1

The IndianFood-7 Dataset was used in this work to increase computer vision algorithms' ability to recognize typical Indian cuisine items. The objective was to create a DL-based detection technique that could recognize these objects.

The study's objective was to train models using the well-chosen dataset using the YOLOv5 and YOLOR algorithms. In particular, the p6 variation of YOLOR and three YOLOv5 variants (YOLOv5s, YOLOv5m, and YOLOv5l) were used. Training was done with a provided image size of 416 by 416 for 150 epochs.

Figures 5.1, 5.2, and 5.3 show the accuracy, F1 score, recall, and mAP evaluation metrics as they change for the YOLOv5s, YOLOv5m, and YOLOv5l models across 150 data points over the periods of training (on the x-axis).

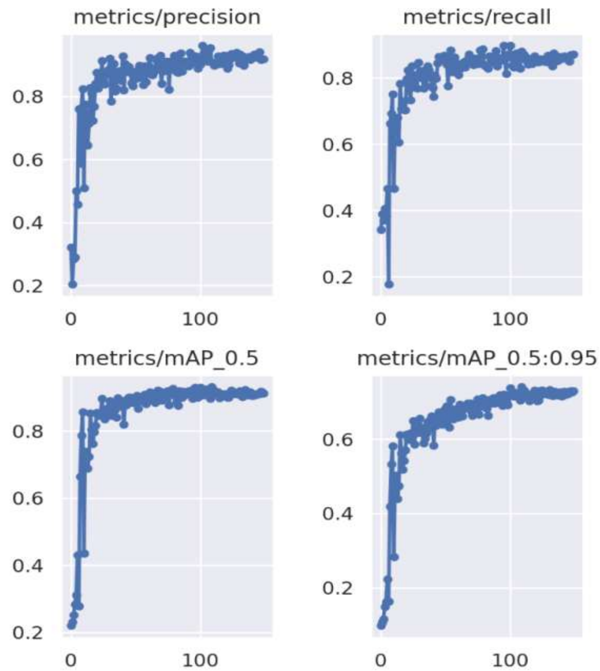


Figure 5.1: Training overview for YOLOv5s beyond 150 epochs

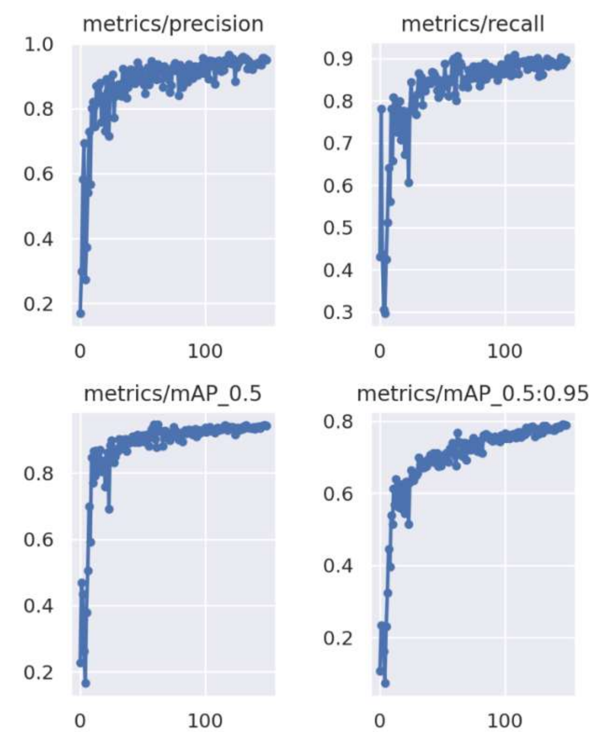


Figure 5.2: Overview of training for YOLOv5m after 150 epochs

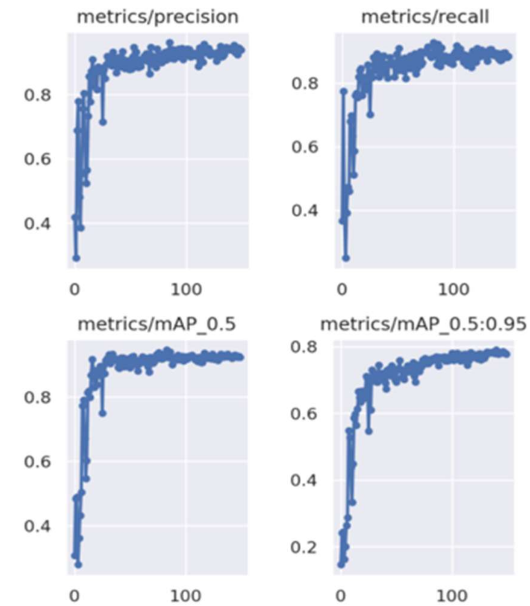


Figure 5.3: Overview of training for YOLOv5l after 150 epochs

The YOLOv5l model's performance in relation to several food categories is shown in Table 5.2. A variety of food classes, including BesanCheela, Idli, Palak Paneer, Dosa, Gulab Jamun, Poha, and Samosa, were assessed throughout the experiment's first phase.

Table 5.2: Thorough examination of the YOLOv51 model for every label

S.No	Classes of Food	Precision	Recall	F1-Score	mAP
1	Dosa	0.790	0.754	0.771	0.740
2	Idli	0.894	0.788	0.838	0.853
3	BesanCheela	0.756	0.818	0.786	0.853
4	Poha	0.894	1.000	0.944	0.995
5	Gulab Jamun	0.955	0.895	0.924	0.930
6	Palak Paneer	0.954	1.000	0.976	0.995
7	Samosa	0.926	0.741	0.823	0.787

The model obtained an accuracy of 0.756, a remember of 0.818, an F1-score of 0.786, and a mAP of 0.853 for the food class labeled "BesanCheela." In a similar vein, the cuisine class designated as "Gulab Jamun" had the following results: 0.955 accuracy, 0.895 recall, 0.924 F1-score, and 0.93 mAP. With an accuracy of 0.881, remember of 0.857, F1-score of 0.869, and mAP of 0.879, the YOLO5l model performed well overall.

Table 5.3: Thorough examination of theYOLOv5m model for every labels

S.No	Classes of Food	Precision	Recall	F1-Score	mAP
1	Dosa	0.820	0.750	0.783	0.745
2	Idli	0.961	0.791	0.868	0.878
3	Samosa	0.909	0.736	0.813	0.807
4	Poha	0.848	1.000	0.918	0.985
5	BesanCheela	0.792	0.691	0.738	0.812
6	Gulab Jamun	0.939	0.868	0.902	0.922
7	Palak Paneer	0.973	1.000	0.986	0.995

The YOLOv5m model's performance in relation to several food categories is manifest in Table 5.3. The result of models for "Dosa" were as follows: mAP of 0.745, F1-score of 0.783, recall of 0.75, and precision of 0.82. In a similar vein, "Idli" had a 0.961 of accuracy, remember of 0.791, F1-score of 0.868, and mAP of 0.878. All things considered, the YOLOv5m model showed 0.892 accuracy, 0.834 recall, 0.862 F1-score, and 0.878 mAP.

Table 5.4: Thorough examination of the YOLOv5s model for each of the labels

S.No	Classes of Food	Precision	Recall	F1-Score	mAP
1	Dosa	0.935	0.721	0.814	0.911
2	Idli	0.942	0.753	0.837	0.833
3	Samosa	0.914	0.778	0.840	0.837
4	Gulab Jamun	0.970	0.848	0.905	0.918
5	Poha	0.930	0.950	0.940	0.987
6	BesanCheela	0.878	0.545	0.672	0.752
7	Palak Paneer	0.970	0.950	0.960	0.993

The performance of the YOLOv5s model is presented in Table 5.4 for a variety of annotated food categories. The model's results for "Palak Paneer" were an F1-score of 0.960, a mAP of 0.993, a remember of 0.95, and a precision of 0.97. In a similar vein, "Samosa" had an accuracy of 0.914, remember of 0.778, F1-score of 0.840, and mAP of 0.837. The YOLOv5s model showed an overall mAP of 0.89, an F1-score of 0.857, a recall of 0.792, and an accuracy of 0.934.

Table 5.5: Using the testing dataset to assess the models

S.No	Model	Precision	Recall	F1-Score	mAP
1	YOLOv5m	0.892	0.834	0.862	0.878
2	YOLOR	0.782	0.899	0.836	0.893
3	YOLOv5l	0.881	0.857	0.869	0.879
4	YOLOv5s	0.934	0.792	0.857	0.890

A thorough examination of the several methods discovered for resolving data collecting is given in Table 5.5. The study revealed that, with a mAP score of 0.893, the YOLOR method performed the best. With an accuracy score of 0.934, the YOLOv5s model activated the finest in terms of accuracy. Furthermore, with an F1-Score of 0.869, the YOLOv5l model showed the best performance in terms of the F1-Score. Significantly higher recall values were found in the bigger methods (YOLOv5l and YOLOR) than in the tiny methods (YOLOv5s and YOLOv5m), suggesting that the larger models are better able to identify Indian cuisine items in images.

Figure 5.4: YOLOv5m (a) dropped Gulab Jamun, which YOLOv5l (b) identified.

Figure 5.5: Idli is distracted by BesanCheela in YOLOv5s (a) revised by YOLOv5l (b)

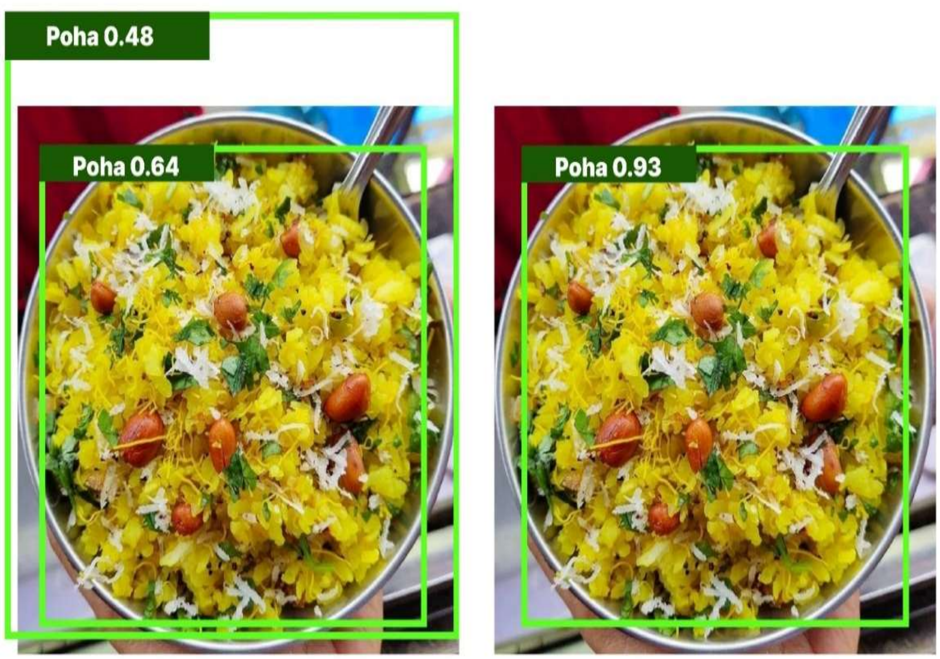


Figure 5.6: YOLOv5s (a) provides more false positives in comparison to YOLOv5l (b)

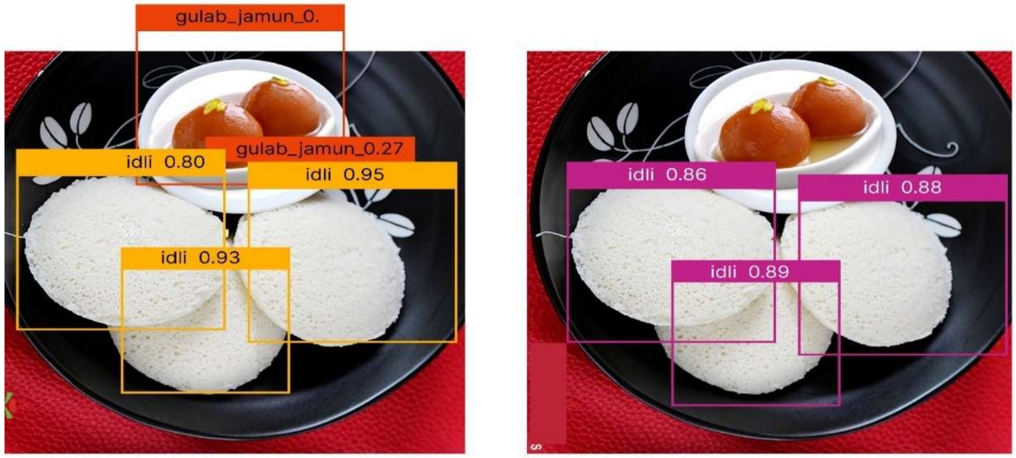


Figure 5.7: YOLOv5l (a) provides more false positives in comparison to YOLOR (b)

Table 5.5 shows that YOLOv5l detects Gulab Jamun, which YOLOv5m missed, because of its greater recall and F1-Score (Figure 5.4). Figure 5.5 shows that YOLOv5s incorrectly recognizes Idli as BesanCheela; however, the bigger YOLOv5l model corrects this problem. As seen in Figure 5.6, YOLOv5s generates more false positives than YOLOv5l. Last but not least, YOLOv5l produces false positives in Figure 5.7, which YOLOR is supposed to correct.

5.5 EXPERIMENTAL RESULTS FOR PHASE 2

In this work, food image identification was studied using CNN, Faster R-CNN, and YOLO as algorithms. A comparative analysis was then conducted. The principal classifier for food images was a CNN, with Faster R-CNN also being employed. In addition, YOLO was included to the food database in order to modify categorization precision. The experiment was carried out using images from the IndianFoodNet dataset that cover a variety of food categories.

The results from the various versions of YOLO—versions 5, 7, and 8—are shown in Tables 5.6, 5.7, and 5.8 accordingly. The comparison is mostly focused on the YOLO5 and YOLO7 versions since Faster R-CNN is thought to perform better than YOLO in segmentation tasks as opposed to object recognition.

Using a model trained to produce the most accurate predictions on the test set, phase 2 of the investigation was conducted. Performance was evaluated using a variety of measures, adding prediction, remember, mAP50, mAP higher than 50, and deduction speed (measured in milliseconds). A total of 1654 images of food from 30 distinct cuisine groups were included in the study.

Table 5.6: YOLO5's Performance in MapReduce, Precision, and Recall

S.No	Class	Images	Precision	Recall	mAP50	mAP 50-95
1	AlooGobhi	41	0.478	0.576	0.545	0.448
2	Aloo Masala	40	0.663	0.603	0.709	0.58
3	Coconut Chutney	58	0.443	0.792	0.632	0.458
4	White Rice	59	0.788	0.790	0.802	0.602
5	Bhatura	88	0.702	0.863	0.906	0.702
6	Biryani	46	0.445	0.779	0.618	0.524
7	BhindiMasala	44	0.612	0.700	0.778	0.639
8	Chai	61	0.660	0.760	0.760	0.445
9	Chole	73	0.578	0.888	0.840	0.646
10	DumAloo	45	0.336	0.260	0.317	0.268
11	Dal	46	0.535	0.557	0.399	0.340
12	Dosa	76	0.756	0.570	0.638	0.386
13	Jalebi	40	0.933	0.924	0.958	0.693
14	GulabJamun	96	0.526	0.919	0.886	0.623
15	Fish Curry	42	0.330	0.245	0.280	0.243
16	Ghevar	47	0.576	0.626	0.632	0.497
17	ShahiPaneer	41	0.605	0.200	0.297	0.255
18	Green Chutney	65	0.788	0.618	0.725	0.495
19	RasMalai	43	0.557	0.416	0.564	0.446
20	Idli	88	0.623	0.976	0.932	0.697
21	Samosa	61	0.695	0.618	0.688	0.446
22	Kebab	88	0.494	0.743	0.674	0.360
23	Poha	36	0.378	0.942	0.880	0.736
24	Kheer	51	0.400	0.502	0.496	0.372
25	Kulfi	68	0.720	0.738	0.788	0.528
26	Lassi	57	0.592	0.912	0.872	0.618
27	Mutton Curry	41	0.133	0.024	0.206	0.173
28	RajmaCurry	43	0.382	0.763	0.528	0.448
29	OnionPakoda	41	0.738	0.976	0.966	0.777
30	PalakPaneer	44	0.657	0.904	0.855	0.705

The YOLO5 model's performance is shown in Table 5.6, with an emphasis on mAP, recall, and accuracy following five epochs.

For example, the model obtained an accuracy of 0.446, a recall of 0.778, a mAP of 0.617, and a mAP higher than 50 of 0.523 for the "Biryani" cuisine class, which had 45 images. In a similar vein, the model achieved an accuracy of 0.556, recall of 0.417, mAP of 0.563, and mAP greater than 50 of 0.445 for the 42 images in the "RasMalai" cuisine class. In summary, the YOLO5 model showed 0.573 accuracy, 0.671 recall, 0.671 mAP50, and 0.504 mAP 50–85.

Table 5.7: YOLO7's Performance in MapReduce, Precision, and Recall

S.No	Class	Images	Precision	Recall	mAP50	mAp50-95
1	AlooGobi	41	2.000	1.000	0.156	0.126
2	Aloo Masala	40	0.346	0.406	0.402	0.326
3	Bhatura	88	0.257	0.817	0.606	0.476
4	BhindiMasala	44	0.256	0.134	0.195	0.163
5	Biryani	46	0.340	0.157	0.170	0.148
6	Chai	61	0.207	0.460	0.268	0.168
7	Chole	73	0.215	0.320	0.193	0.152
8	Coconut Chutney	58	0.249	0.422	0.193	0.140
9	Dal	46	0.218	0.700	0.389	0.318
10	Dosa	76	0.197	0.454	0.215	0.137
11	DumAloo	45	0.129	0.073	0.138	0.118
12	Fish Curry	42	0.038	0.024	0.075	0.059
13	Ghevar	47	2.000	1.000	0.051	0.037
14	Green Chutney	65	0.477	0.683	0.629	0.473
15	GulabJamun	96	0.365	0.413	0.334	0.198
16	Idli	88	0.316	0.702	0.563	0.430
17	Jalebi	40	0.566	0.203	0.368	0.268
18	Kebab	88	0.350	0.552	0.460	0.209
19	Kulfi	68	0.137	0.608	0.216	0.142
20	Kheer	51	0.115	0.350	0.123	0.0893
21	Lassi	57	0.300	0.733	0.367	0.218
22	Mutton Curry	41	0.053	0.076	0.074	0.054
23	OnionPakoda	41	0.276	0.574	0.273	0.300
24	PalakPaneer	44	0.205	0.934	0.585	0.494
25	Poha	36	0.276	0.372	0.234	0.192

S.No	Class	Images	Precision	Recall	mAP50	mAp50-95
26	RasMalai	43	0.206	0.763	0.418	0.328
27	RajmaCurry	43	0.223	0.385	0.249	0.206
28	Samosa	61	0.217	0.566	0.237	0.115
29	ShahiPaneer	41	0.197	0.26	0.138	0.13
30	White Rice	59	0.378	0.662	0.436	0.296

The YOLO7 model's performance is shown in Table 5.7, with an emphasis on mAP, recall, and accuracy. For example, the model obtained an accuracy of 0.377, a recall of 0.661, a mAP of 0.435, and a mAP higher than 50 of 0.295 for the food class "White Rice," which included 59 images. Similarly, the model achieved an accuracy of 0.275, recall of 0.575, mAP of 0.272, and mAP greater than 50 of 0.2 for the food class "Onion Pakoda," which consisted of 40 images. The YOLO7 model showed an overall mAP50 of 0.291, a precision of 0.299, a recall of 0.422, and a mAP 50-85 of 0.212.

Table 5.8: YOLO8's Performance in MapReduce, Precision, and Recall

S.No	Class	Images	Precision	Recall	mAP50	mAp50-95
1	AlooGobi	41	0.806	0.516	0.780	0.607
2	Aloo Masala	40	0.618	0.822	0.818	0.678
3	Bhatura	88	0.932	0.642	0.856	0.698
4	BhindiMasala	44	0.739	0.866	0.908	0.725
5	Biryani	46	0.840	0.823	0.907	0.741
6	Chai	61	0.675	0.660	0.698	0.465
7	Chole	73	0.763	0.902	0.926	0.727
8	Coconut Chutney	58	0.927	0.657	0.872	0.728
9	Dal	46	0.894	0.745	0.886	0.748
10	Dosa	76	0.714	0.508	0.602	0.373
11	DumAloo	45	0.875	0.478	0.747	0.607
12	FishCurry	42	0.682	0.469	0.663	0.517
13	Ghevar	47	0.802	0.687	0.798	0.586
14	GreenChutney	65	0.765	0.757	0.798	0.627
15	GulabJamun	96	0.734	0.821	0.856	0.628
16	Idli	88	0.832	0.852	0.876	0.688
17	Jalebi	40	0.98	0.975	0.995	0.765
18	Kebab	88	0.655	0.515	0.574	0.363
19	Kheer	68	0.447	0.800	0.702	0.563
20	Kulfi	51	0.785	0.639	0.757	0.525
21	Lassi	57	0.734	0.626	0.768	0.588

S.No	Class	Images	Precision	Recall	mAP50	mAp50-95
22	MuttonCurry	41	0.789	0.576	0.806	0.596
23	OnionPakoda	41	0.945	0.760	0.918	0.732
24	PalakPaneer	44	0.950	0.889	0.947	0.788
25	Poha	36	0.735	2.000	0.948	0.757
26	RasMalai	43	0.801	0.596	0.775	0.598
27	RajmaCurry	43	0.816	0.820	0.877	0.748
28	Samosa	61	0.695	0.620	0.676	0.468
29	ShahiPaneer	41	0.520	0.740	0.665	0.559
30	WhiteRice	59	0.829	0.736	0.842	0.634

Table 5.8 presents the YOLO8 model's performance with an emphasis on mAP, recall, and precision. For example, the model obtained an accuracy of 0.927, a recall of 0.657, a mAP of 0.872, and a mAP higher than 50 of 0.728 in the "Coconut Chutney" cuisine class using 58 images. Comparably, the template attained a level of precision of 0.785, a recall of 0.639, a mAP of 0.757, and a mAP greater than 50 of 0.525 for the "Kulfi" food class, which included 51 images. The YOLO8 model showed an overall mAP 50 of 0.807, mAP 50-85 of 0.628, recall of 0.719, and accuracy of 0.775. According to these findings, YOLO performs better in terms of categorization accuracy than other combinations.

These measurements shed light on how well the YOLO8 algorithm recognizes and categorizes different kinds of food.

Table 5.9: Assessment of YOLO variations' performance

S.No	Model	Images	Precision	Recall	mAP50	mAP 50-95
1	YOLO8	1654	0.775	0.719	0.807	0.628
2	YOLO5	1654	0.573	0.671	0.671	0.504
3	YOLO7	1654	0.299	0.422	0.291	0.212

Three YOLO models are shown in detail in Table 5.9. YOLO5 obtained a 0.573 accuracy value, a 0.671 recall value, a 0.671 mAP50 value, and a 0.504 mAP 50–95 value. YOLO7 achieved a mAP 50 value of 0.291, a mAP 50-95 value of 0.212, a recall value of 0.422, and a precision score of 0.299. In the meantime, YOLO8 showed a mAP

50 value of 0.807, a mAP 50-95 value of 0.628, a precision value of 0.775, and a recall value of 0.719. Notably, compared to YOLO7, which had poorer accuracy at epoch 5, the models trained on YOLO5 and YOLO8 had noticeably superior results with greater recall, mAP, and precision values at epoch 5.

Selecting the deployment model is essential for object detection. More accurate prediction models are typically slower during inference, and vice versa. Regarding the IndianFoodNet dataset, the findings for YOLO7 differ, while the inference of training loss is low and metrics accuracy and recall are following up in YOLO5 and YOLO8. According to the YOLO training summary at five epochs, YOLO8's accuracy is 0.775 times greater than YOLO7's and YOLO5's. When compared to YOLO5 (recall value = 0.671) and YOLO8 (recall value = 0.719), YOLO7 has the lowest recall value.

The training summary for the three YOLO5 models, YOLO7 models, and YOLO8 models is shown in Figures 5.8, 5.9, and 5.10, in that order.

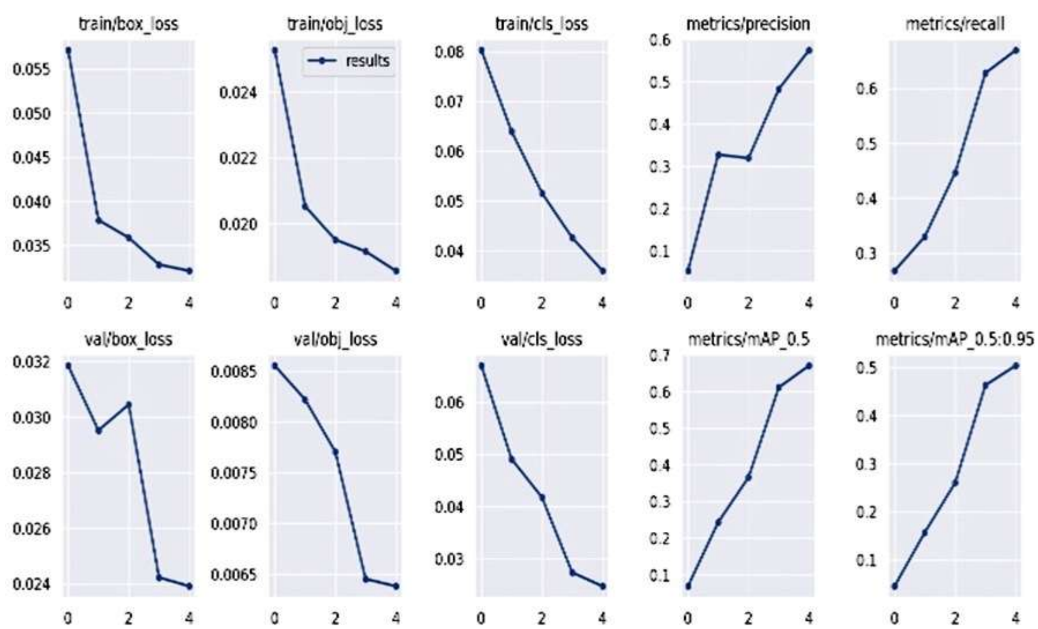


Figure 5.8: YOLO5 training overview across five epochs

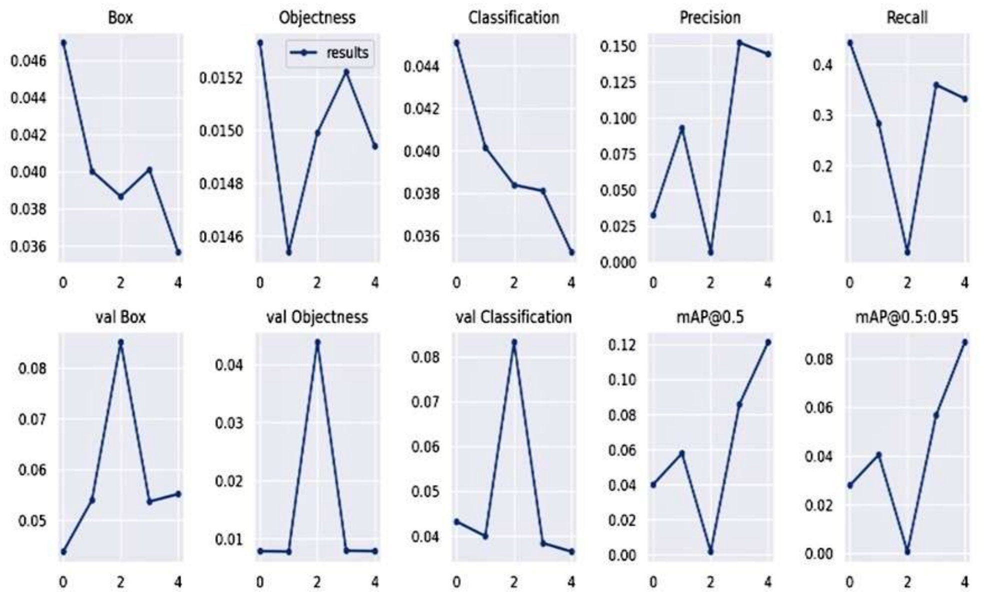


Figure 5.9: YOLO7 training overview across five epochs

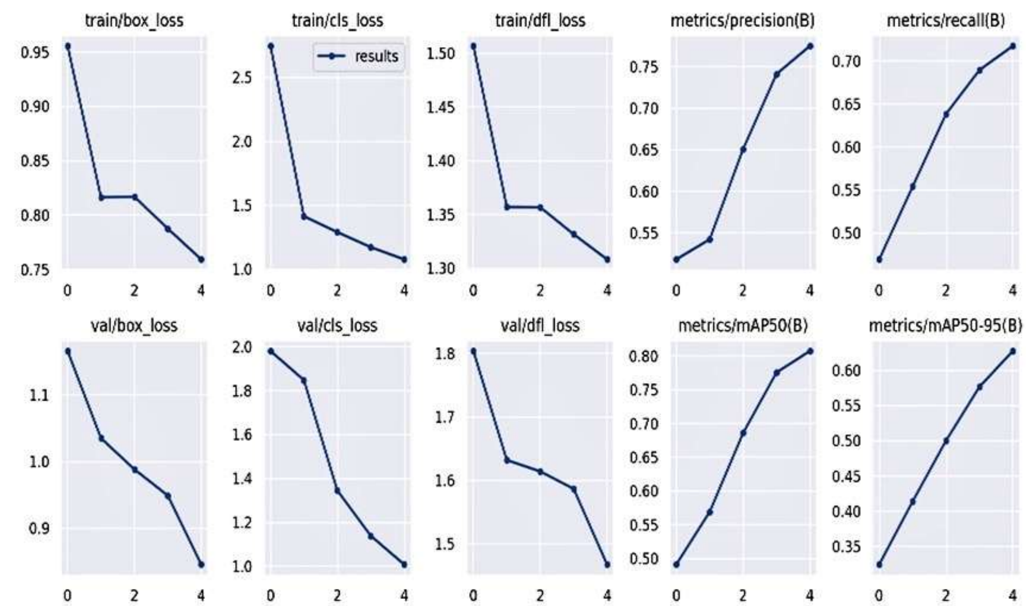


Figure 5.10: YOLO8 training overview across five epochs

The forecasts of the YOLO7 model are shown to be less accurate than those of the YOLO8 and YOLO5 models. YOLO7 receives a mAP score of.291, whereas YOLO8 and YOLO5 have mAP ratings of.807 and.671, respectively.

5.6 EXPERIMENTAL RESULTS FOR PHASE 3

This section looks at how well the suggested model performs. Finding meal items and estimating their energy level are the goals of this endeavour. For this reason, a hybrid framework incorporating DL algorithms was created. Three various images were used in this investigation; they were segmented using MaskR-CNN and categorized using Yolov5. Compared to other approaches, this strategy produced the better levels of metrics, "precision, specificity, negative predictive value (NPV), recall, Matthews correlation coefficient (MCC), F-score, sensitivity, false positive rate (FPR), false discovery rate (FDR), and false negative rate (FNR)".

Table 5.10: Images of suggested and existent objects, segmented and detected techniques

Method	Input Image	Segmented Image	Detected Image
Mask (R-CNN) + YOLO V5			
			
			
Mask (R-CNN)			

Method	Input Image	Segmented Image	Detected Image
			
			
YOLO			
			
			
CNN			
			

Method	Input Image	Segmented Image	Detected Image
			

The segmented and recognized images obtained using the suggested methodology as well as the current methodologies are shown in Table 5.10. In particular, the combination of YOLO V5 and MaskR-CNN is compared against both YOLO and CNN techniques, as well as Mask R-CNN alone.

Table 5.11: Volume estimation and calorie estimation

S.NO	Food Items	Volume Estimation (cubic cm)	Calories Estimation(calories)
1	Gulab Jamun	40015	300
2	Fish Curry	47944	217
3	Rajma Curry	28245	207

The estimated volume and calorie content of different foods are shown in Table 5.11. For instance, it is estimated that 217 calories are found in fish curry, 207 calories are found in Rajma curry, and 300 calories are found in gulab jamun. These approximations shed light on the nutritional makeup of the many foods examined in the investigation.

5.6.1 Performance analysis

The suggested method's effectiveness is evaluated in terms of Fscore, MCC, NPV, FPR, FDR, FNR, sensitivity, accuracy, specificity, recall, and precision. Additionally, the accuracy of the proposed model is contrasted with the outcomes of multiple categorization models, including CNN, YOLO, and Mask (R-CNN). The following section covers the complete performance of the recognised images using various categorization models and other metrics.

Table 5.12: Comparison of the Current Techniques with Proposed Method

S.No	Metrics of Evaluation	Mask (R-CNN) and YOLO V5	Mask (R-CNN)	YOLO	CNN
1	Accuracy	0.97135	0.96235	0.95135	0.92135
2	Precision	0.96435	0.95025	0.94135	0.93435
3	Sensitivity	0.96764	0.95364	0.94764	0.93764
4	Fscore	0.96574	0.95174	0.94443	0.93594
5	NPV	0.94582	0.93882	0.92682	0.91582
6	FNR	0.01710	0.02410	0.03310	0.04510
7	MCC	0.96231	0.95031	0.94351	0.93231
8	FPR	0.01640	0.02340	0.02840	0.03440
9	Specificity	0.96462	0.95162	0.94262	0.93442
10	FDR	0.01580	0.01880	0.02390	0.03480

The assessment measurements of the suggested Mask (R-CNN) +YOLO V5 are contrasted with those of alternative methods of categorization in Table 5.12 above. When compared to Mask (R-CNN), YOLO, and CNN, Mask (R-CNN) +YOLO V5 had the highest accuracy of 0.97126 across the different categorization models. In comparison to Mask (R-CNN) +YOLO V5, the average precision is 0.92126 for CNN, 0.96226 for Mask (R-CNN), and 0.95126 for YOLO. Our suggested technique has greater sensitivity and specificity than existing models, at 0.96754 and 0.96452, respectively. The MCC and NPV of our suggested approach were 0.96241 and 0.94592, respectively. Our suggested technique has greater accuracy and f-score (0.96415 and 0.96584, respectively) another method. The overview examination of the suggested and finalizing methods are shown in Figure 5.11.

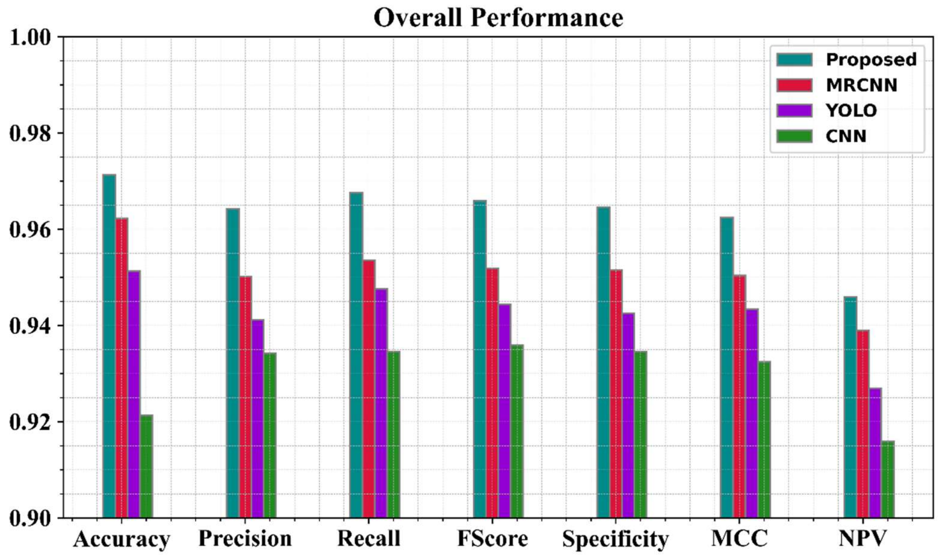


Figure 5.11: Comparative Analysis of Proposed and Current Approaches

Table 5.12 compares the new method error metrics with those of the present approaches. It proves that, analysed another current approach, our proposal has a very small error value.0.0161 for FNR,0.0154 for FPR, and 0.0148 for FDR, for example. A comparison of error metrics across various methods is visible in Figure 5.12.

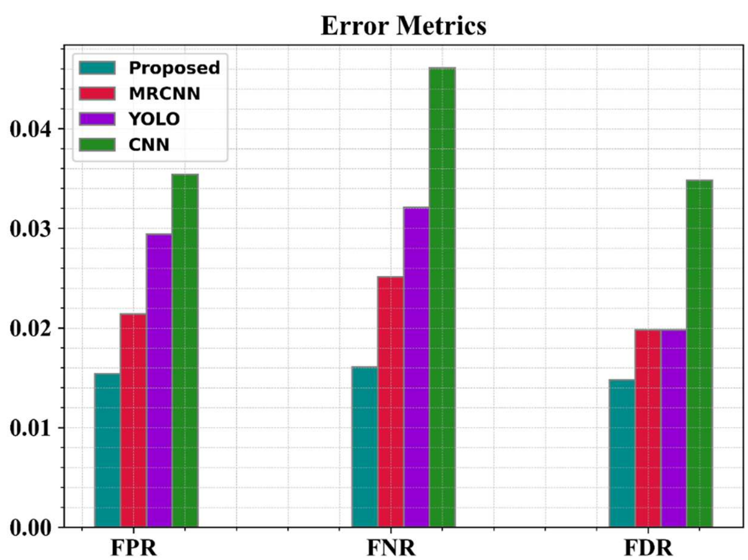


Figure 5.12: Comparative analysis of different approaches' error metrics

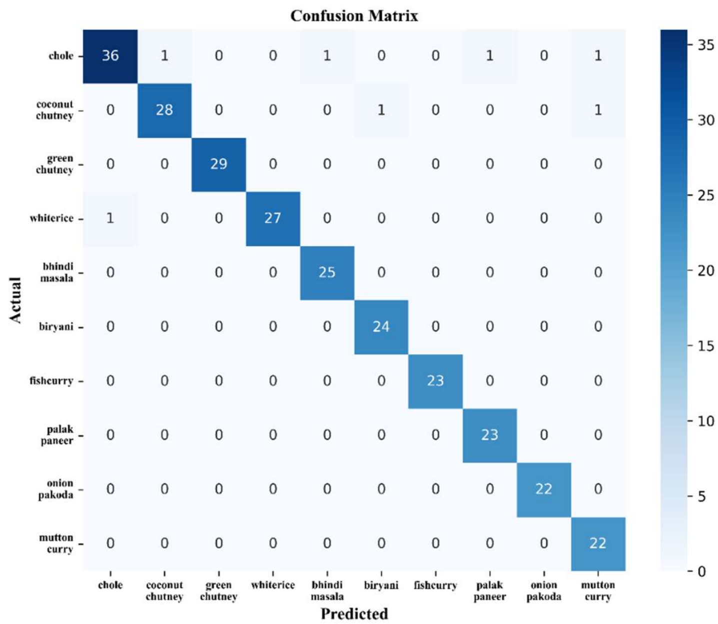


Figure 5.13: The Confusion matrix for proposed method

The confusion matrix for the proposed technique is visible in Figure 5.13 as well as the confusion matrix for the approaches that are currently in use, such as CNN in Figure 5.16, YOLO in Figure 5.15, and Mask (R-CNN) in Figure 5.14.

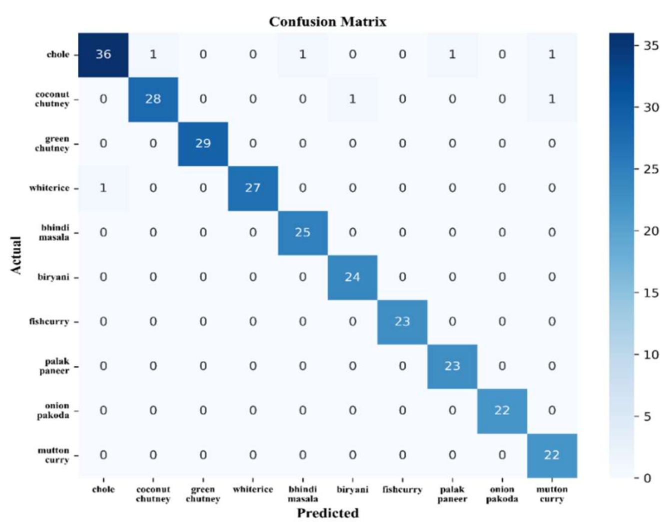


Figure 5.14: Matrix of confusion for Mask (R-CNN)

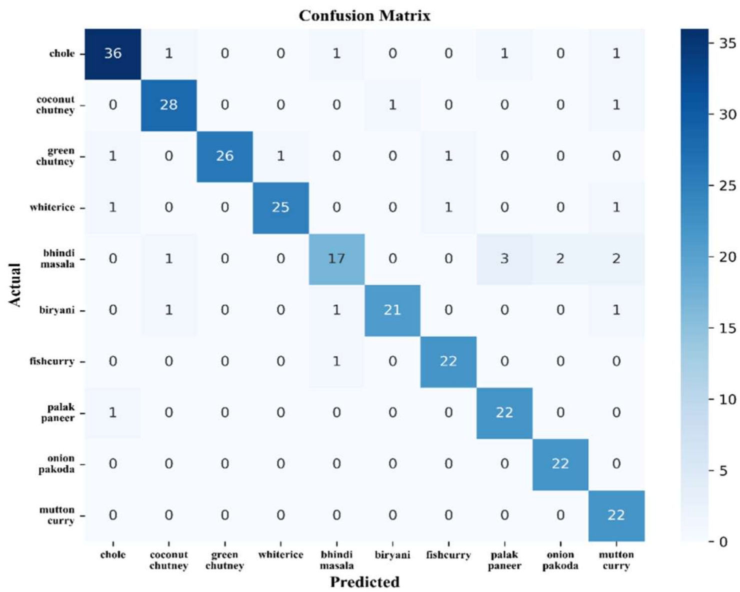


Figure 5.15: Matrix of confusion for YOLO

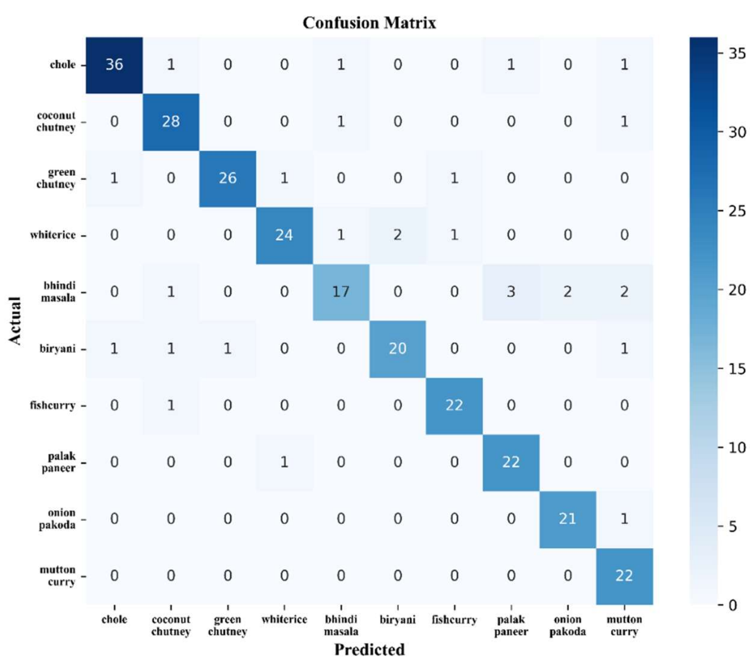


Figure 5.16: Matrix of confusion for CNN

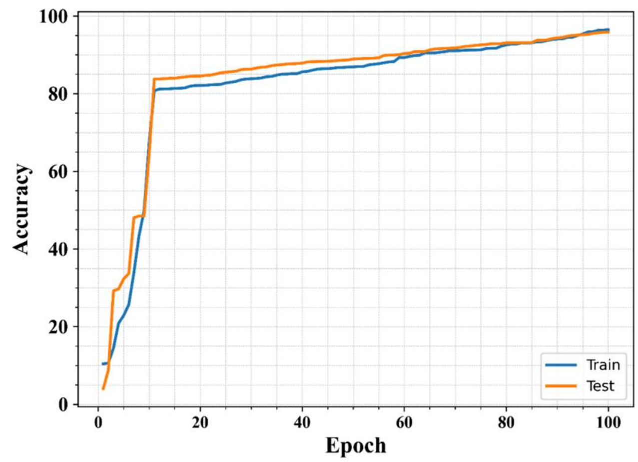


Figure 5.17: Accuracy-Epoch graph

A graph showing the link between accuracy and epoch during the course of training is shown in Figure 5.17. The precision of the method rises with the quantity of epochs, suggesting that the model's capacity to accurately categorize or predict outcomes has improved. The increasing accuracy trend indicates that the method is picking up different talents from the learning set and improving with time.

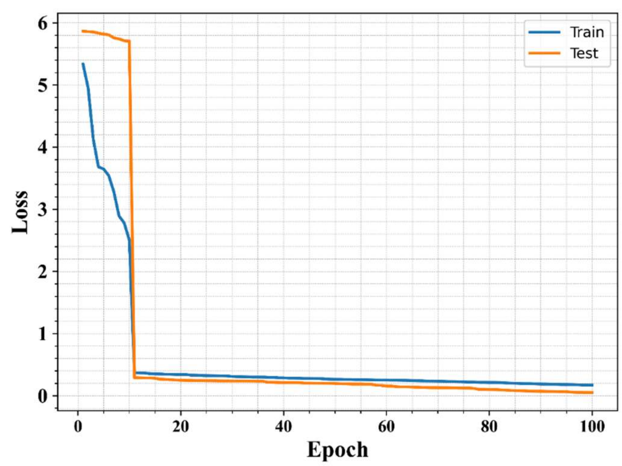


Figure 5.18: Loss-Epoch graph

A graph showing the link between loss and epoch during the model's training process is manifest in Figure 5.18. The loss value drops as the quantity of epochs rises, showing that the methods performance is steadily improving. The model is effectively learning from the training data and improving prediction accuracy, as seen by this decreased trend in loss.

5.6.2 Statistical Analysis for Proposed Model

Statistical analysis is done to verify that the proposed model is comparatively more efficient than other models that have been researched. The output data, which included accuracy, precision, FNR, etc., selected the t-test. It is a statistical method for the comparison of the means of two groups. It is employed in hypothesis testing for determining whether two groups are distinct. And also, whether a technique or treatment actually impacts the population of interest. Table 5.13 provides t-test for Accuracy and shows the result of proposed model more efficient than comparison with existing model. The real mean, standard deviation, and confidence interval values are also obtained by the performed T-test in addition to the t values.

Table 5.13: t-Test for Accuracy

	Proposed	**MRCNN**	**YOLO**	**CNN**
Actual Mean	0.97104	0.96098	0.95070	0.92064
No. of Values (n)	10	10	10	10
t-value	5730.8	3604.8	7677.6	8238.1
Degrees of Freedom	9	9	9	9
p-value (2-tailed) ("< e-4")	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.0005358275	0.0008430105	0.000391578	0.0003533962
Variance	2.871111e-07	7.106667e-07	1.533333e-07	1.248889e-07
Median	0.9711	0.9609	0.9507	0.9206
95% Confidence Interval	0.9706567-0.9714233	0.9603769-0.9615831	0.9504199-0.9509801	0.9203872-0.9208928

Similarly, Table 5.14 for precision, Table 5.15 for sensitivity, Table 5.16 for fmeasure, Table 5.17 for specificity, shows the result of proposed model more efficient than comparison with existing model.

Table 5.14: t-Test for Precision

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.96247	0.95005	0.94040	0.93221
No. of Values (n)	10	10	10	10
t-value	1942.2	57003	11330	2939.9
Degrees of Freedom	9	9	9	9
p-value (2-tailed) (“< e-4”)	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.001567057	5.270463e-05	0.0002624669	0.001002719
Variance	2.455667e-06	2.777778e-09	6.888889e-08	1.005444e-06
Median	0.9633	0.9500	0.9405	0.9323
95% Confidence Interval	0.961349-0.963591	0.9500123-0.9500877	0.9402122-0.9405878	0.9314927-0.9329273

Table 5.15: t-Test for Sensitivity

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.96486	0.95326	0.94460	0.93539
No. of Values (n)	10	10	10	10
t-value	2238.3	17601	1939	2160.6
Degrees of Freedom	9	9	9	9
p-value (2-tailed) (“< e-4”)	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.001363166	0.0001712698	0.001540563	0.001369063
Variance	1.858222e-06	2.933333e-08	2.373333e-06	1.874333e-06
Median	0.9646	0.9533	0.9442	0.9357
95% Confidence Interval	0.9638848-0.9658352	0.9531375-0.9533825	0.9434979-0.9457021	0.9344106-0.9363694

Table 5.16: t-Test for FMeasure

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.96364	0.95166	0.94249	0.93379
No. of Values (n)	10	10	10	10
t-value	2476.8	43040	3729.1	3966.1
Degrees of Freedom	9	9	9	9
p-value (2-tailed) (“< e-4”)	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.001230357	6.992059e-05	0.0007992357	0.0007445356
Variance	1.513778e-06	4.888889e-09	6.387778e-07	5.543333e-07
Median	0.9637	0.9516	0.9422	0.9339
95% Confidence Interval	0.9627599-0.9645201	0.95161- 0.95171	0.9419183-0.9430617	0.9332574-0.9343226

Table 5.17: t-Test for Specificity

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.96255	0.95078	0.94142	0.93186
No. of Values (n)	10	10	10	10
t-value	3195.7	5473.1	3420.9	2365.7
Degrees of Freedom	9	9	9	9
p-value (2-tailed) (“< e-4”)	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.0009524821	0.000549343	0.000870249	0.001245615
Variance	9.072222e-07	3.017778e-07	7.573333e-07	1.551556e-06
Median	0.9628	0.9507	0.9414	0.9317
95% Confidence Interval	0.9618686-0.9632314	0.950387-0.951173	0.9407975-0.9420425	0.9309689-0.9327511

Table 5.18 for FPR, Table 5.19 for FNR, Table 5.20 for FDR provides t-test and shows the result of proposed model decreases in comparison with existing model as measurement of the percentage of positive cases that a test misidentified or mistakenly classified as positive.

Table 5.18: t-Test for FPR

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.01513	0.02008	0.02751	0.03526
No. of Values (n)	10	10	10	10
t-value	412.64	106.89	73.15	1037.3
Degrees of Freedom	9	9	9	9
p-value (2-tailed) ("< e-4")	2.2e-16	2.786e-15	8.432e-14	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.0001159502	0.0005940445	0.001189257	0.0001074968
Variance	1.344444e-08	3.528889e-07	1.414333e-06	1.155556e-08
Median	0.01530	0.02005	0.02745	0.03520
95% Confidence Interval	0.01504705-0.01521295	0.01965505-0.02050495	0.02665926-0.02836074	0.0351831-0.0353369

Table 5.19: t-Test for FNR

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.01599	0.0218	0.02817	0.04602
No. of Values (n)	10	10	10	10
t-value	685.29	76.026	51.699	2301
Degrees of Freedom	9	9	9	9
p-value (2-tailed) ("< e-4")	2.2e-16	5.963e-14	1.904e-12	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	7.378648e-05	0.0009067647	0.001723079	6.324555e-05
Variance	5.444444e-09	8.222222e-07	2.969e-06	4e-09
Median	0.01600	0.02200	0.02765	0.04600
95% Confidence Interval	0.01593722-0.01604278	0.02115134-0.02244866	0.02693738-0.02940262	0.04597476-0.04606524

Table 5.20: t-Test for FDR

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.01426	0.01798	0.02254	0.03353
No. of Values (n)	10	10	10	10
t-value	384.18	93.671	74.205	88.561
Degrees of Freedom	9	9	9	9
p-value (2-tailed) (“< e-4”)	2.2e-16	9.133e-15	7.415e-14	1.512e-14
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.0001173788	0.0006069962	0.0009605554	0.001197265
Variance	1.377778e-08	3.684444e-07	9.226667e-07	1.433444e-06
Median	0.01420	0.01795	0.02300	0.03365
95% Confidence Interval	0.01417603-0.01434397	0.01754578-0.01841422	0.02185286-0.02322714	0.03267353-0.03438647

Similarly, Table 5.21, Table 5.22 provides t-test for NPV, MCC and shows the result of proposed model more efficient than comparison with existing model.

Table 5.21: t-Test for NPV

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.95322	0.93392	0.92244	0.91235
No. of Values (n)	10	10	10	10
t-value	764.78	1681.1	1427.8	1550.9
Degrees of Freedom	9	9	9	9
p-value (2-tailed) (“< e-4”)	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.003941461	0.001756765	0.002042983	0.001860257
Variance	1.553511e-05	3.086222e-06	4.173778e-06	3.460556e-06
Median	0.9526	0.9345	0.9230	0.9116
95% Confidence Interval	0.9504004-0.9560396	0.9326633-0.9351767	0.9209785-0.9239015	0.9110193-0.9136807

Table 5.22: t-Test for MCC

	Proposed	MRCNN	YOLO	CNN
Actual Mean	0.96076	0.95017	0.94144	0.93121
No. of Values (n)	10	10	10	10
t-value	4841.6	22465	3164.8	4169.6
Degrees of Freedom	9	9	9	9
p-value (2-tailed) (“< e-4”)	2.2e-16	2.2e-16	2.2e-16	2.2e-16
Significant (alpha=0.05)?	Yes	Yes	Yes	Yes
SD	0.0006275172	0.0001337494	0.0009406853	0.000706242
Variance	3.937778e-07	1.788889e-08	8.848889e-07	4.987778e-07
Median	0.9607	0.9502	0.9411	0.9311
95% Confidence Interval	0.9603111-0.9612089	0.9500743-0.9502657	0.9407671-0.9421129	0.9307048 0.9317152

5. 7 STRENGTHS OF THE MODEL

1.High Accuracy in Food Segmentation and Classification: The use of Mask R-CNN for segmentation and YOLO V5 for classification allows the model to accurately identify and isolate individual food items, even in complex scenes. This combination minimizes errors that could arise from overlapping or partially obscured food items. The model’s ability to segment and classify each food item accurately is crucial for calorie estimation, as misclassification or incomplete segmentation could lead to incorrect calorie counts.

2.Efficiency and Real-Time Capability: YOLO V5 is known for its speed and efficiency, which makes this model suitable for real-time applications, such as mobile health apps and wearable devices. The model’s efficiency means it could potentially be deployed on devices with limited computational resources, allowing for wide accessibility. This efficiency also enables quicker processing times, which is essential for applications requiring instant feedback on calorie content.

3. Scalability and Flexibility: The model can be expanded to recognize more food items by training on a larger, more diverse dataset. This scalability is beneficial for a range of applications, from general dietary tracking to specialized uses, like tracking particular food groups or regional cuisines. The approach is also flexible enough to incorporate future improvements, such as the integration of new, more advanced deep learning architectures.

4.Holistic Approach to Calorie Estimation: By combining segmentation, classification, and volume estimation, the model provides a comprehensive and context-aware calorie estimation. This approach considers not only the type of food but also the quantity, which can result in more accurate calorie calculations compared to traditional image recognition models that might only classify the food type. This approach is also suitable for diverse dietary scenarios, as it considers both calorie density and food portion sizes, making it adaptable to different dietary needs.

5.Usefulness in Health and Nutrition Applications: The model has significant potential for applications in health and nutrition, such as dietary tracking, calorie management, and promoting healthy eating habits. This can support individuals managing conditions like diabetes, obesity, or cardiovascular diseases by helping them monitor their daily calorie intake. Additionally, the model could be integrated into apps that assist people with specific dietary restrictions or requirements, providing tailored information about the nutritional content of foods.

5.8 LIMITATION AND WEAKNESS OF THE MODEL

1.Dependence on Dataset Quality and Diversity: The model's performance is highly dependent on the quality, diversity, and representativeness of the training dataset. If the dataset lacks certain types of foods or only represents food in specific contexts (e.g., plated in a particular way), the model may struggle to generalize to new food presentations, cultural cuisines, or uncommon food items. This limitation could lead to inaccurate calorie estimations when the model encounters foods outside of its training data, potentially reducing its reliability in real-world settings.

2.Challenges in Volume Estimation: Accurately estimating food volume from 2D images is inherently challenging, especially when there is no depth information. Without additional cues (like a reference object of known size), the model may struggle to precisely measure the dimensions and volume of irregularly shaped food items, leading to inaccurate calorie calculations. Volume estimation also becomes challenging for foods that have varying shapes and sizes (e.g., soups, salads), as these items don't conform to standard volume formulas.

3.Sensitivity to Image Quality and Lighting Conditions: The model may be sensitive to image quality and lighting variations, which can impact the accuracy of both segmentation and classification. Poor lighting or low-resolution images can lead to missed or misclassified food items, affecting the overall calorie estimate. This sensitivity limits the model's usability in environments with inconsistent lighting, such as restaurants or outdoor settings, unless pre-processing techniques are applied to standardize image quality.

4.Potential for Misclassification in Mixed Dishes: Mixed dishes (e.g., casseroles, salads, sandwiches) that contain multiple ingredients can be difficult for the model to classify accurately. YOLO V5 may struggle to differentiate individual components within a mixed dish, which can lead to inaccurate classification and calorie estimation. This limitation is particularly problematic for dishes with hidden ingredients, where visual cues alone may not be sufficient to determine the food's calorie content.

5.Lack of Nutritional Context Beyond Calories: While the model estimates calorie content, it doesn't provide a complete nutritional profile, such as macronutrient breakdown (carbohydrates, proteins, fats) or micronutrient content (vitamins, minerals). In certain applications, calorie count alone may be insufficient, as users might also need information about the overall nutritional quality of their diet. This limitation restricts the model's utility for users who are tracking specific nutrients, not just calories, which could be an area for future enhancement.

6.Limited Adaptability to Portion Sizes and Consumption Variability: People often consume foods in varying portion sizes, and it may be difficult for the model to account for partially eaten items. For instance, a half-eaten sandwich or a partially consumed plate may lead to overestimations of calorie intake if the model cannot adjust for these variations. This limitation is especially relevant in real-world usage, as the actual portion consumed might not match the estimated volume, leading to inaccuracies.

The model performs well in controlled settings and with simple, well-defined foods, but it struggles with complex, mixed, or irregularly shaped dishes, as well as foods that are difficult to measure accurately from images alone. Improvements in dataset diversity, more sophisticated segmentation techniques, or depth estimation methods could help address some of these challenges and enhance the model's robustness in real-world applications.

5.9 SUMMARY

In this chapter, the focus was on setting up simulations and analyzing the performance of three crucial contributions. The initial contribution introduced a IndianFood-7 dataset to improve computer vision-based methods' ability to recognize common Indian foods. The aim of this study is to train the YOLOv5 and YOLOR algorithms on a curated dataset in order to recognize Indian culinary items. Three YOLOv5 versions—YOLOv5s, YOLOv5m, and YOLOv5l—as well as the p6 form of YOLOR are used in the training. Using a Google Colab Notebook, training was done over 150 epochs with the entering image width set at 416 x 416. After examination, the results showed it performed the fine, with a mAP score of 0.893. With a score of 0.869, its fared improve in terms of the F1-Score measure. Notably, the larger models (YOLOv5l and YOLOR) exhibited higher recall values, indicating their superior ability to accurately detect Indian food items compared to smaller models (YOLOv5m, YOLOv5s).

The second component of the work uses a DL method to teach a method for the identification of calories in various meal products. Roboflow's concept was used to create a Google Colab notebook for testing and training purposes. Notebooks from the original repository connected to several variations of YOLO were also included. Because of the combined high accuracy and speed of the YOLO5, YOLO7, and YOLO8 algorithms, this tailored dataset makes training and comparing them easier. It was discovered that, in comparison to YOLO8 and YOLO5, the predictions made by the YOLO7 model were less accurate. YOLO7 had a score of.291, whereas YOLO8 and YOLO5 received mAP ratings of.807 and.671, respectively. The findings showed that all algorithms performed over the threshold of 0.80 mAP with sufficient accuracy. With a mAP of 0.67, YOLO5 performed the fine in terms of precision, and YOLO8 had the quickest inference time (5 ms).

A hybrid framework utilizing DL algorithms to assess food products' calorie content has been presented in the third part of the study. Python was used to implement the suggested method on a Windows PC with a 1.6 GHz Intel Core i5 processor and 4GB of RAM. The success of the model was examined and contrasted to other sorting models, like Mask (R-CNN), CNN, and YOLO, with an array of actions such as FNR, MCC, sensitivity, recall, specificity, NPV, FDR, FPR, and accuracy. The suggested technique outperformed other models with precision and sensitivity of 0.96452and 0.96754, respectively. The suggested approach produced NPV and MCC values of 0.96241 and 0.94592, respectively. Its accuracy and F-score were greater than those of the other three models, at 0.96415 and 0.96584, respectively. The assessment metrics suggest that the suggested approach shows noticeably reduced error levels in contrast to the strategies that are currently in use. Comparing the suggested strategy to other models, it shows an overall high accuracy of 97.12% for a single image (taken randomly). When we take 10% of testing images in a loop randomly, then 94% accuracy is achieved, which is high overall, providing a robust and reliable solution for identifying and estimating the content of calorie in Indian food items.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 INTRODUCTION

The chapter wraps up the research on "Recognition of Food Type Using Deep Learning and Calorie Estimation" and wraps up the thesis. It concludes with closing thoughts and summarizes the primary findings of this thesis, along with recommendations for additional research avenues.

6.2 CONCLUSION

This study aimed at identifying food items and estimating their calorie content to promote health maintenance. It made three significant contributions. First, the creation of the IndianFood-7 Dataset was instrumental in advancing computer vision techniques for recognizing common Indian dishes. Various strategies were employed to bolster the model's robustness. The goal was to develop a deep learning-based system capable of recognizing Indian cuisine accurately. To achieve this, YOLOv5 and YOLOR algorithms were applied, with different versions of YOLOv5 (s, m, l) and the p6 variant of YOLOR being trained on the dataset. Performance metrics such as recall, F1-score, accuracy, and mean Average Precision (mAP) were used to evaluate these models' effectiveness. Notably, larger models like YOLOv5l and YOLOR demonstrated higher recall values compared to smaller ones (YOLOv5s and YOLOv5m), indicating their superior capability in detecting Indian food items more precisely.

In the second contribution, the study aimed to determine the caloric content of various food items using a DL algorithm and to train a model to recognize them. Previous studies couldn't find a suitable image dataset for the research requirements, so a custom dataset was created. Initially, a list of commonly consumed food items in different regions of India was compiled. After that, a wide range of food-related images were gathered, and each food item was found inside the images using bounding boxes. Training and testing of the YOLO5, YOLO7, and YOLO8 algorithms—known for their rapidity and accuracy—were made possible by this unique set. The accuracy of all algorithms was demonstrated to be good by the results, with a mAP above 0.80. The accuracy of predictions from YOLO8 and YOLO5 was higher than that of YOLO7. With a mAP of 0.67, YOLO5 had the highest accuracy, while YOLO8 had the quickest inference time (5 ms).

The study's final contribution suggested a hybrid approach for evaluating food products' calorie content that uses DL algorithms. Initially, mask (R-CNN) was used to segment the images. After segmentation, features were extracted from a segmented images and food items were classified using the framework of YOLOV5. Following the identification of the food items, their dimensions were determined, and these approximations were used to compute the volume of each food item. The number of calories was then calculated based on the food's volume. The models above were made using the food photos found in the dataset. Its efficacy was assessed against various classification models as “CNN, YOLO, and Mask (R-CNN)”, using a variety of metrics as “speed, sensitivity, specificity, precision, recall, F-score, MCC, NPV, FPR, FDR, and FNR”. According to these assessment measures, the suggested approach had substantially lower error values than the already used techniques. Comparing the suggested strategy to other models, it showed an overall high accuracy of 97.12%.

6.3 FUTURE SCOPE

To put it briefly, the thesis presents a new approach to calorie calculation and food identification. The results of this study, which uses DL for food item detection, are a first step in developing a robust detection model. This approach will help people identify food items more precisely since measurements were used to determine the most accurate variation. Furthermore, this work allows computer vision specialists to improve current techniques and leverage food-related visual signals more effectively. Even if the proposed architecture is extensive, it may be improved.

In future studies,

- Expanding the Indian Food Dataset is vital for identifying popular Indian dishes. More images of existing food items should be added to boost the accuracy of object detection models, along with including new food labels.

- Estimating calorie intake using advanced YOLO variants like YOLO5 or YOLO8 shows promise. However, significant hurdles are to overcome, particularly in assessing food volume.
- Evaluating 3D vision algorithms for real-world for food object analysis [122].
- Determining portion sizes based on volume is crucial for promoting healthy eating habits. However, processed or strangely shaped meals may be difficult for DL models to consume. For these models, producing accurate training data is expensive and time-consuming.
- Models might not translate well to novel meals or changes in presentation, which would limit their applicability.
- Improving the variety of photos and points of view in each scenario, applying, and refining the artificial dataset to more nearly resemble photos taken in real life [123].
- Accuracy is also affected by the backdrop and lighting conditions. To overcome these obstacles and provide accurate meal volume prediction in practical settings, more study, improved data gathering techniques, and the creation of more reliable DL models will be needed.
- It will incorporate a wider variety of food shapes and container designs, enhancing the system's ability to recognize and compute the range of food shapes found worldwide and establishing a platform for testing models of international cuisine.

REFERENCES

1. “Wikipedia” [Online] Available https://en.wikipedia.org/wiki/Obesity#cite_note-WHO2015-1.(Date:15jan24, Time:11:05pm)

2. The American Medical Association Annual Meeting (June 18, “No Title.” [Online]) Available http://www.ama-assn.org/ama/pub/news/news/2013/2013-06-18-new-ama-policies-annual-meeting.page. .(Date:15jan24, Time:11:15pm)

3. F. Ofli, Y Aytar, I. Weber, R. Al Hammouri, and A. Torralba (2017). Is Saki # delicious?: The food perception gap on instagram and its relation to health, *In International Conference on World Wide Web* (pp.509–518). DOI:10.1145/3038912.3052663

4. L. Bossard, M. Guillaumin, and L. Van Gool (2014). Food-101-mining discriminative components with random forests, *In European Conference on Computer Vision(*pp.446-461). DOI: 10.1007/978-3-319-10599-4_29

5. A. Culotta (2014). Estimating County health statistics with twitter. *In ACM Conference on Human Factors in Computing Systems(*pp.1335–1344). https://doi.org/10.1145/2556288.2557139

6. S. Abbar, Y. Mejova, and I. Weber (2015). You tweet what you eat: Studying food consumption through twitter *Proceedings of the ACM Conference on Human Factors in Computing Systems(*pp.3197–3206). https://doi.org/10.1145/2702123.2702153

7. K. Mccrickerd and C. G. Forde (2016). Sensory influences on food intake control: Moving beyond palatability. *Obesity Reviews (Wiley Online Library)* Volume 17*,* Issue 1*(*pp.18–29). https://doi.org/10.1111/obr.12340

8. P. Achananuparp, E. Lim, and V. Abhishek (2018). Does journaling encourage healthier choices? Analyzing healthy eating behaviors of food journalers. *In The International Digital Health Conference.* https://arxiv.org/pdf/1805.01129

9. F. Cordeiro, E. Bales, E. Cherry, and J. Fogarty (2015). Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. *In Proceedings of the ACM Conference on Human Factors in Computing Systems.* (pp. 3207–3216). https://doi.org/10.1145/2702123.2702154

10. D. A. Williamson, H. R. Allen, P. Davis Martin, A. Alfonso, B. Gerald, and A. Hunt (2004). Digital photography: A new method for estimating food intake in cafeteria settings. *Eating and Weight Disorders Ewd 9, 1* (pp. 24–28) https://link.springer.com/article/10.1007/BF03325041

11. MouzhiGe, M. Elahi, I Fernandez-Tobias, F. Ricci, and D. Massimo (2015). Using tags and latent factors in a food recommender system. *In Proceedings of the 5th International Conference on Digital Health (*pp.105–112). DOI:10.1145/2750511.2750528

12. AronCulotta (2014). Estimating county health statistics with twitter. *In ACM Conference on Human Factors in Computing Systems. (*pp.1335–1344). https://doi.org/10.1145/2556288.2557139

13. D. Fried, M. Surdeanu, S. Kobourov, and M. Hingle (2014). Analyzing the language of food on social media. *In IEEE International Conference on Big Data*. (pp778–783). DOI: 10.1109/BigData.2014.7004305

14. "Obesity", nhs.uk(2018). *[Online]. Available*: https://www.nhs.uk/conditions/obesity/ [Accessed: 22- Feb- 2018].

15. "Defining Childhood Obesity | Overweight & Obesity | CDC"(2018), *Cdc.gov, [Online]*. Available: https://www.cdc.gov/obesity/childhood/ defining. html. [Accessed: 22- Feb- 2018].

16. J. Steinberger and S. R. Daniels (2003), "Obesity, Insulin Resistance, Diabetes, and Cardiovascular Risk in Children," *AHAIASA Journal Circulation, vol. 107, no. 10(*pp. 1448–1453). https://doi.org/10.1161/01.CIR.0000060923.07573.F2

17. J. Sorof and S. Daniels (2002), "Obesity hypertension in children: A problem of epidemic proportions," *AHAIASA Journal Hypertension, vol. 40, no. 4.* (pp. 441–447). https://doi.org/10.1161/01.HYP.0000032940.33466.12

18. S. Mayor (2016), "Over a third of children aged 10-11 in England are overweight or obese," *BMJ (Online), vol. 355*. https://pubmed.ncbi.nlm.nih.gov/27815251/

19. AH Mokdad, JS Marks, DF Stroup, JL Gerberding(2004). Actual causes of death in the United States, 2000. *JAMA 291*:*1238–45*, DOI: 10.1001/jama.291.10.1238.

20. "Reducing obesity: future choices - GOV.UK"(2018), *Gov.uk, 2018. [Online]. Available:*https://www.gov.uk/government/publications/reducing-obesity-future-choices. [Accessed: 22- Feb- 2018].

21. S. Bleich, D. Cutler, C. Murray, and A. Adams (2007), "Why is the developed world obese?" *National Bureau of Economic Research, Tech. Rep*. DOI:10.1146/annurev.publhealth.29.020907.090954

22. W. H. O. (WHO), "Global strategy on diet, physical activity and health, fivety seventh world health assembly resolution," *Tech. Rep*. https://www.who.int/publications/i /item/9241592222.

23. W. H. Organization (2003), "Diet, nutrition and the prevention of chronic diseases," *WHO Technical Report Series*, vol. 916.

24. S. Baxter, W. Thompson, M. Litaker, F. Frye, and C. Guinn (2002), "Low accuracy and low consistency of fourth-graders school breakfast and school lunch recalls," *Journal of the American Dietetic Association*, vol. 102, no. 3(pp. 386–395).

25. J. Trabulsi and D. Schoeller (2001), "Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 281, no. 5, (pp. E891–E899). https://doi.org/10.1152/ajpendo.2001.281.5.E891

26. L. E. Burke, J. Wang, and M. A. Sevick (2011), "Self-Monitoring in Weight Loss: A Systematic Review of the Literature," *J. Am. Diet. Assoc., vol. 111, no. 1(*pp. 92–102). DOI: 10.1016/j.jada.2010.10.008

27. J. P. Chaput, L. Klingenberg, A. Astrup, and A. M. Sjödin (2011), "Modern sedentary activities promote overconsumption of food in our current obesogenic environment," *Obes. Rev., vol. 12*, (no. 501). DOI: 10.1007/s13679-012-0026-7

28. M. C. Carter, V. J. Burley, C. Nykjaer, and J. E. Cade (2013), "Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial.," *J. Med. Internet Res., vol. 15*, (no. 4).

29. X.Ding et al. (2019), "Reading Face, Reading Health: Exploring Face Reading Technologies for Everyday Health" *in Proceedings of the 33rd Annual ACM*

Conference on Human Factors in Computing Systems - CHI '15, 2015(pp.1-13). https://doi.org/10.1145/3290605.33004

30. M. J. Hutchesson, M. E. Rollo, R. Callister, and C. E. Collins (2015), "Self-Monitoring of Dietary Intake by Young Women: Online Food Records Completed on Computer or Smartphone Are as Accurate as Paper-Based Food Records but More Acceptable," *J. Acad. Nutr. Diet., vol. 115, no. 1(*pp. 87–94). https://doi.org/10.1016/j.jand.2014.07.036

31. C.-F. Chung, E. Agapie, J. Schroeder, S. Mishra, J. Fogarty, and S. A. Munson (2017), "When Personal Tracking Becomes Social: Examining the Use of Instagram for Healthy Eating," (pp. 1674–1687). doi: 10.1145/3025453.3025747

32. J. M. Vaterlaus, E. V. Patten, C. Roche, and J. A. Young (2015), "Gettinghealthy: The perceived influence of social media on young adult health behaviors," (pp. 151–157). DOI:10.1016/j.chb.2014.12.013

33. F. Jimoh et al. (2018), "Comparing Diet and Exercise Monitoring Using Smartphone App and Paper Diary: A Two-Phase Intervention Study," *JMIR mHealth uHealth, vol. 6, no. 1,* (p. e17). doi:10.2196/mhealth.7702

34. Parappurathu S., Kumar A., Bantilan M.C.S., & Joshi P.K. (2015): "Food Consumption Pattern and Dietary Diversity in Eastern India: evidence from village level studies". *Food Security. 7(5),* 1031-1042.

35. Hu F.B. (2002): "Dietary pattern analysis: a new direction in nutritional epidemiology", *Current Opinion in Lipidology,* 13, 3-9. DOI: 10.1097/00041433-200202000-00002.

36. J.Wardle ,et.al (2012): "Parental Feeding Style and the Inter-generational Transmission of Obesity Risk", *Obesity Research, Wiley online library*(pp 453-462). https://doi.org/ 10. 1038 /oby. 2002.63

37. A. Mary, and M.T. Ruel (2004). "Dietary Diversity Is Associated with Child Nutritional Status: Evidence from 11 Demographic and Health Surveys". *Journal of Nutrition*, 134(10), 2579–2585. https://doi.org/10.1093/jn/134.10.2579

38. Esfarjani F., Roustaee R., Nasrabadi F. M. Ahmad Esmaillzadeh (2013): "Major Dietary Patterns in Relation to Stunting among Children in Tehran, Iran", *Journal of Health, Population and Nutrition, 2013 Jun; 31(2): 202-210.*

39. W. Jensen, M., Afzal, S., Jess, T., Nordestgaard, B. G., & Allin, K. H(2020). Body mass index and risk of infections: a Mendelian randomization study of 101,447 individuals. *European journal of epidemiology, 35(4)*(pp 347-354) DOI: 10.1007/s10654-020-00630-7

40. D. G. Lowe (2004), Distinctive image features from scale-invariant keypoints. *Intl. Journal of Computer Vision*. DOI:10.1023/B:VISI.0000029664.99615.94

41. Food Surveys Research Group Beltsville, MD: Agricultural Research Service. USDA Food and Nutrient Database for Dietary Studies, 3.0, 2008.

42. F. Zhu, A. Mariappan, C. Boushey, D. Kerr, K. Lutes, D. Ebert, and E. Delp (2008), "Technology-assisted dietary assessment," in *Proceedings of SPIE*, vol. 6814. NIH Public Access, (p. 681411).

43. K. Kitamura, C. de Silva, T. Yamasaki, and K. Aizawa (2010), "Image processing-based approach to food balance analysis for personal food logging," in *Multimedia and Expo (ICME), 2010 IEEE International Conference on*. IEEE, (pp. 625– 630).

44. J. Tufano and B. Karras (2005), "Mobile ehealth interventions for obesity: a timely opportunity to leverage convergence trends," *Journal of medical Internet research*, vol. 7, no. 5. doi: 10.2196/jmir.7.5.e58

45. M. Sun, Q. Liu, K. Schmidt, J. Yang, N. Yao, J. D. Fernstrom, M. H. Fernstrom, J. P. DeLany, and R. J. Sclabassi (2008), "Determination of food portion size by image processing," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. (pp. 871–874).

46. R. Sclabassi, M. Sun *et al. (2007)*, "A video-based algorithm for food intake estimation in the study of obesity," in *Bioengineering Conference, 2007. NEBC'07. IEEE 33rd Annual Northeast*. (pp. 298–299).

47. M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney (2009), "Recognition and volume estimation of food intake using a mobile device," in *Applications of Computer Vision (WACV), 2009 Workshop on*. IEEE, (pp. 1–8).

48. S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar(2010), "Food recognition using statistics of pairwise local features," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, (pp. 2249–2256).

49. F. Zhu, M. Bosch, T. Schap, N. Khanna, D. S. Ebert, C. J. Boushey, and E. J. Delp (2011), "Segmentation assisted food classification for dietary assessment," in *Proceedings of SPIE*, vol. 7873. NIH Public Access, p. 78730B.

50. M. Bosch, F. Zhu, N. Khanna, C. J. Boushey, and E. J. Delp (2011), "Combining global and local features for food identification in dietary assessment," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE(pp. 1789–1792).

51. F. Kong and J. Tan (2011), "Dietcam: Regular shape food recognition with a camera phone," in *Body Sensor Networks (BSN), 2011 International Conference on*. IEEE (pp. 127–132).

52. Y. Saraf (2006), "Algorithms for image segmentation," Ph.D. dissertation, Birla Institute of Technology and Science.

53. Im Cho, Y., & Palvanov, A. (2019). A New Machine Learning Algorithm for Weather Visibility and Food Recognition. *J. Robotics Netw. Artif. Life*, *6*(1), 12-17. DOI:10.2991/jrnal.k.190531.003

54. Savakar, D. G., & Anami, B. S. (2009). Recognition and classification of food grains, fruits and flowers using machine vision. *International Journal of Food Engineering*, *5*(4). DOI:10.2202/1556-3758.1673

55. Schroeder, V., Evans, E. D., Wu, Y. C. M., Voll, C. C. A., McDonald, B. R., Savagatrup, S., & Swager, T. M. (2019). Chemiresistive sensor array and machine learning classification of food. *ACS sensors*, *4*(8), 2101-2108. DOI: 10.1021/acssensors.9b00825

56. Magnus, I., Virte, M., Thienpont, H., & Smeesters, L. (2021). Combining optical spectroscopy and machine learning to improve food classification. *Food Control*, *130*, 108342. DOI:10.1016/j.foodcont.2021.108342

57. Sharif, A., Abbasi, Q. H., Arshad, K., Ansari, S., Ali, M. Z., Kaur, J., ... & Imran, M. A. (2021). Machine learning enabled food contamination detection using RFID and internet of things system. *Journal of Sensor and Actuator Networks*, *10*(4), 63. https://doi.org/10.3390/jsan10040063

58. Mandal, B., Puhan, N. B., & Verma, A. (2018). Deep convolutional generative adversarial network-based food recognition using partially labeled data. *IEEE Sensors Letters*, *3*(2), (1-4). DOI:10.1109/LSENS.2018.2886427

59. Wei, P., & Wang, B. (2022). Food image classification and image retrieval based on visual features and machine learning. *Multimedia Systems*, *28*(6), (2053-2064.) DOI:10.1007/s00530-020-00673-6

60. Fakhrou, A., Kunhoth, J., & Al Maadeed, S. (2021). Smartphone-based food recognition system using multiple deep CNN models. *Multimedia Tools and Applications*, *80*(21-23), (33011-33032). DOI:10.1007/s11042-021-11329-6

61. McAllister, Patrick, et al(2018). "Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets." *Computers in biology and medicine* 95 (217-233).

62. Wang, Y., Wu, J., Deng, H., & Zeng, X. (2021). Food image recognition and food safety detection method based on deep learning. *Computational Intelligence and Neuroscience*. doi: 10.1155/2021/1268453

63. Pouladzadeh, P., & Shirmohammadi, S. (2017). Mobile multi-food recognition using deep learning. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *13*(3s), (1-21). https://doi.org/ 10.1145/3063592

64. Liu, Y. C., Onthoni, D. D., Mohapatra, S., Irianti, D., & Sahoo, P. K. (2022). Deep-Learning-Assisted Multi-Dish Food Recognition Application for Dietary Intake Reporting. *Electronics*, *11*(10), 1626. https://doi.org/10.3390/ elect ronics11101626

65. Mezgec, S., Eftimov, T., Bucher, T., & Seljak, B. K. (2019). Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment. *Public health nutrition*, *22*(7), (1193-1202). doi: 10.1017/S1368980018000708

66. Mezgec, S., & Koroušić Seljak, B. (2017). NutriNet: a deep learning food and drink image recognition system for dietary assessment. *Nutrients*, *9*(7), 657. DOI: 10.3390/nu9070657

67. Ciocca, Gianluigi, G Micali, and P Napoletano(2020). "State recognition of food images using deep features." *IEEE Access* 8: 32003-32017. DOI:10.1109/ ACCESS.2020.2973704

68. Liu, Chang, et al. (2017) "A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure." *IEEE Transactions on Services Computing* 11.2 (249-261).

69. Yu, Qian, Dongyuan Mao, and Jingfan Wang (2016). "Deep learning-based food recognition." *Technical report, Stanford University*.

70. R. Agarwal, T. Choudhury, NJ Ahuja & T. Sarkar (2023). Hybrid Deep Learning Algorithm-Based Food Recognition and Calorie Estimation. *Journal of Food Processing and Preservation. DOI:10.1155/2023/6612302*

71. Haque, Rakib Ul, et al. (2022): "Lightweight and Parameter-Optimized Real-Time Food Calorie Estimation from Images Using CNN-Based Approach." *Applied Sciences* https://doi.org /10.3390/ app12199733

72. Dai, Y., Park, S., & Lee, K. (2022). Utilizing Mask R-CNN for Solid-Volume Food Instance Segmentation and Calorie Estimation. *Applied Sciences*, *12*(21), 10938. https://doi.org/10.3390/app122110938

73. Mashraf, Chowdhury Z, et al.(2021) *FoodieCal: a convolutional neural network based food detection and calorie estimation system*. Diss. Brac University, DOI: 10.1109/NCCC49330.2021.9428820

74. Ege, T., & Yanai, K. (2018). Image-based food calorie estimation using recipe information. *IEICE TRANSACTIONS on Information and Systems*, *101*(5), (1333-1341). DOI:10.1587/transinf.2017MVP0027

75. Emmanuel, W. S., & Minija, S. J. (2018). Fuzzy clustering and Whale-based neural network to food recognition and calorie estimation for daily dietary assessment. *Sādhanā*, *43*, (1-19). DOI:10.1007/s12046-018-0865-3

76. Wasif, S Mohd, et al. (2019) "Food calorie estimation using machine learning and image processing." *International Journal of Advanced Research, Ideas and Innovations in Technology* 5.2: (1627-1630). https://www.ijariit.com/manuscripts/v5i2/V5I2-1935.pdf

77. Puli, Mrs Srilatha, et al. (2023)"Food Calorie Estimation Using Convolutional Neural Network." *Journal of Survey in Fisheries Sciences* 2665-2671.

78. Lee, K. S. (2023). Multispectral Food Classification and Caloric Estimation Using CNN *Foods*, *12*(17), 3212. DOI: 10.3390/foods12173212

79. Kumar, R. Dinesh, et al. (2021) "Recognition of food type and calorie estimation using neural network." *The Journal of Supercomputing* 1-22. DOI:10.1007/s11227-021-03622-w

80. Liang, Hong, et al. (2020) "CEP: calories estimation from food photos." *International Journal of Computers and Applications* 42.6): 569-577. DOI:10.1080/1206212X.2018.1486558

81. Deng, Dongping, et al. (2023)"Defect Detection in Food Using Multispectral and High-Definition Imaging Combined with a Newly Developed Deep Learning Model." *Processes* 11.12: 3295. DOI:10.3390/pr11123295

82. Ulrici, Alessandro, et al. (2012) "Automated identification and visualization of food defects using RGB imaging: Application to the detection of red skin defect of raw hams." *Innovative Food Science & Emerging Technologies* 16: 417-426.

83. Wang, C., & Xiao, Z. (2021). Potato surface defect detection based on deep transfer learning. *Agriculture*, *11*(9), 863. https://doi.org/ 10.3390/ agriculture 11090863

84. Pham, V. H., & Lee, B. R. (2015). An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm. *Vietnam Journal of Computer Science*, *2*, 25-33. DOI 10.1007/s40595-014-0028-3

85. Da Costa, A. Z., Figueroa, H. E., & Fracarolli, J. A. (2020). Computer vision based detection of external defects on tomatoes using deep learning. *Biosystems Engineering*, *190*, 131-144. DOI:10.1016/j.biosystemseng.2019.12.003

86. Siddiqi, R. (2019). Automated apple defect detection using state-of-the-art object detection techniques. *SN Applied Sciences*, *1*(11), 1345. DOI:10.1007/s42452-019-1393-4

87. Zhang, B., Huang, W., Gong, L., Li, J., Zhao, C., Liu, C., & Huang, D. (2015). Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier. *Journal of Food Engineering*, *146*, 143-151. DOI:10.1016/j.jfoodeng.2014.08.024

88. A AlMughrabi, A Galán, R Marques, P. Radeva, (2024)" FoodMem: Near Real-time and Precise Food Video Segmentation", https://arxiv.org/abs/2407.12121.

89. Cao, J Zhang, Hao Shi, Kunyu Peng, Y Zhang, Hui Zhang, R Stiefelhagen, K Yang, (2024)" Occlusion-Aware Seamless Segmentation", https://arxiv.org/pdf/2407.02182.

90. Z. -Q. Zhao, P. Zheng, S. -T. Xu and X. Wu, (2019) "Object Detection with Deep. Learning: A Review," *in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212-3232,* doi: 10.1109/TNNLS. 2018.2876865.

91. Cheng, H., Jiang, X., Sun, Y., & Wang, J. (2001). Color image segmentation: advances and prospects. *Pattern Recognition, 34(12), 2259-2281*. DOI:10.1016/S0031-3203(00)00149-7

92. Tompson, J., Goroshin, R., Jain, A., LeCun, Y., &Bregler, C. (2015). Efficient Object Localization Using Convolutional Networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

93. REDMON, J. – FARHADI, A. (2017). YOLO9000: Better, faster, stronger. In Proceedings of the *2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 6517–6525*. DOI: https://doi.org/10.1109/cvpr.2017.690

94. DU, J. 2018. Understanding of object detection based on CNN family and YOLO. *In Journal of Physics: Conference Series, vol. 1004, article no. 012029.* DOI: https://doi.org/10.1088/1742-6596/1004/1/012029

95. NELSON, J. – SOLAWETZ, J. (2020). YOLOv5 is here: State-of-the-art object detection at 140 FPS. *Available at:* https://blog.roboflow.com/yolov5-is-here/

96. Wang, Y.; Cao, T.; Yang, J.; Zheng, Y.; Fang, Z.; Deng, X.; Wu, J. (2021) Camouflaged Object Detection Based on Improved YOLO v5 Algorithm. *Comput. Sci. 48*, 226–232. (In Chinese) doi: 10.11896/jsjkx.210100058

97. Yan, B., Fan, P., Lei, X., Liu, Z., Yang, F., (2021). A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sensing*, 13(9), 1619. https://doi.org/10.3390/rs13091619

98. X. Zhu, S. Lyu, X. Wang, and Q. Zhao (2021), ''TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, pp. 2778–2788.5 https://arxiv.org/abs/2108.11539

99. WANG, C. Y. – YEH, I. H. – LIAO, H. Y. M. (2021). You only learn one representation: Unified network for multiple tasks. *arXiv preprint arXiv:2105.04206.* DOI: https://doi.org/10.48550/arXiv.2105.04206

100. Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Ma, Y. (2016). Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. *In Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14, pp. 37-48*. https://doi.org/10.1007/978-3-319-39601-9_4

101. Nelis, J.L.D., Tsagkaris, A.S., Dillon, M.J., Hajsolva, J., Elliott, C.T. (2020). Smartphone-based optical assays in the food safety field. *TrAC Trends in Analytical chemistry, 129: 115934.* https://doi.org/10.1016/j.trac.2020.115934

102. Bhargava, A., Bansal, A. (2021). Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University-Computer and Information Sciences, 33(3): 243-257.* https://doi.org/10.1016/j.jksuci.2018.06.002

103. https://blog.roboflow.com/whats-new-in-yolov8/#what-is-yolov8.(Date:3oct23, Time:9:05pm)

104. https://learnopencv.com/ultralytics-yolov8/#YOLOv8-vsYOLOv5(Date:3oct23, Time:9:15pm)

105. https://github.com/ultralytics/ultralytics/issues/189(Date:8nov23, Time:9:25pm)

106. World Health Organization. (2011) Obesity Study. *[Online]*. http://www. who. int /mediacentre/factsheets/fs311/en/index.html (Date:8nov23, Time:9:35pm)

107. World Health Organization. (2012) World Health Statistics. *[Online]*. http:// www.who.int/gho/publications/world_health_statistics/2012/en/index.html (Date:8nov23, Time:9:40pm)

108. Pouladzadeh, P., Kuhad, P., Peddi, S.V.B., Yassine, A. and Shirmohammadi, S. (2016), May. Food calorie measurement using deep learning neural network. *In 2016 IEEE international instrumentation and measurement technology conference proceedings (pp. 1-6). IEEE.* DOI: 10.1109/I2MTC.2016.7520547

109. Ren S, He K, Girshick R, Sun J. (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems 2015; 39:91–9.* https://arxiv.org/abs/1506.01497

110. J. Wu (2017), "Introduction to Convolutional Neural Networks," *National Key Lab for Novel Software Technology*.

111. I. W. S. E.P, A. Y. Wijaya and R. Soelaiman (2016), "Klasifikasi Citra Menggunakan Convolutional Neural Networks (CNN) pada Caltech 101," *Journal Teknik ITS,* vol. 5

112. F. H. B. E. Z. Haiqiang (2017), "Combining Convolutional and Recurrent Neural Networks for Human Skin Detection," *IEEE Signal Processing,* vol. 24.

113. Wesley Tay 1 , Bhupinder Kaur 1 , Rina Quek 1 , Joseph Lim 1 and Christiani Jeyakumar Henry 1,2 (2020),"Current Developments in Digital Quantitative Volume Estimation for the Optimisation of Dietary Assessment", *Nutrients 2020, 12, 1167;* doi:10.3390/nu12041167 www.mdpi.com/journal/nutrients.

114. Wenxin Yin, Xizhe Zang, Lei Wu, Xuehe Zhang and Jie Zhao (2024)," A Distortion Correction Method Based on Actual Camera Imaging Principles",

Sensors 2024, 24, 2406. https://doi.org/10.3390/s24082406 https://www.mdpi.com/ journal/sensors.

115. B. Huang and S. Zou, (2022)" A New Camera Calibration Technique for Serious Distortion", *Processes 2022, 10, 488*. https://doi.org/ 10.3390/pr10030488 https://www.mdpi.com/journal/processes.

116. Burke, C.J., Aleo, P.D., Chen, Y.C., Liu, X., Peterson, J.R., Sembroski, G.H. and Lin, J.Y.Y. (2019). Deblending and classifying astronomical sources with Mask R-CNN deep learning. *Monthly Notices of the Royal Astronomical Society*, *490*(3), pp.3952-3965. https://arxiv.org/abs/1908.02748

117. Karatoprak, E. and Seker, S., (2019). An improved empirical mode decomposition method using variable window median filter for early fault detection in electric motors. *Mathematical problems in engineering*.

118. Zhang, J., Song, X., Feng, J. and Fei, J., 2021. X-Ray image recognition based on improved Mask R-CNN algorithm. *Mathematical Problems in Engineering*, *2021*, pp.1-14.

119. X. Zheng, C. Liu, Y. Gong, Q. Yin, W. Jia and M. Sun (2023)," Food volume estimation by multi-layer superpixel", *Mathematical Biosciences and Engineering Volume 20, Issue 4, 6294–6311*. DOI: 10.3934 /mbe.2023271

120. W. Tay 1, Bhupinder Kaur 1, Rina Quek 1, Joseph Lim 1 and Christiani Jeyakumar Henry (2020) ," *Current Developments in Digital Quantitative Volume Estimation for the Optimisation of Dietary Assessment", Nutrients 2020, 12, 1167*; doi:10.3390/nu12041167 www.mdpi.com/journal/nutrients.

121. P. by Blogger, Indian food calorie chart, *URL* http://indianfood recipeswith pictures.blogspot.com/.

122. Y. Chen, et.al," MetaFood3D: Large 3D Food Object Dataset with Nutrition Values", *[892]* arXiv:2409.01966

123. C. Tai, et.al" NutritionVerse: Empirical Study of Various Dietary Intake Estimation Approaches", *[1190]* arXiv:2309.07704 (replaced)

PUBLISHED PAPERS

1. R. Agarwal, T. Choudhury, NJ. Ahuja, T. Sarkar (2023)," Hybrid Deep Learning Algorithm- Based Food Recognition and Calorie Estimation", *Journal of Food Processing and Preservation (Hindawi)volume 2023, issue 1. https://doi.org/10.1155/2023/6612302*

2. R. Agarwal, T. Choudhury, NJ. Ahuja, T. Sarkar (2023)," IndianFoodNet: Detecting Indian Food Items Using Deep Learning",*International Journal of Computational Methods and Experimental Measurements(IJCMEM),page 221-232*. https:// doi.org/10.18280 /ijcmem.110403

3. R. Agarwal, N. Bansal,T. Choudhury, NJ. Ahuja, T. Sarkar (2023)," IndianFood-7: Detecting Indian Food Items Using Deep Learning-Based Computer Vision", *International Conference in Advances and Applications of Artificial Intelligence and Machine Learning(ICAAAIML-2022) (LNEE,volume 1078)* https://link.springer.com/chapter/10.1007/978-981-99-5974-7_2

4. R. Agarwal, NJ. Ahuja (2024)," Counting Calories with Deep Learning: A New Frontier in Nutrition", *International Conference on Optimization Techniques in Engineering and Technology Engineering (ICOTET 2024)[Paper Presented and Accepted]*

5. Dataset Link(https://universe.roboflow.com/indianfoodnet/indianfoodnet)

plag.docx

by V R C

Submission date: 13-Nov-2024 05:02PM (UTC+0700)
Submission ID: 2514281575
File name: plag.docx (10.72M)
Word count: 28317
Character count: 160023

plag.docx

ORIGINALITY REPORT

6%	3%	5%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	downloads.hindawi.com Internet Source	1%
2	"Advances and Applications of Artificial Intelligence & Machine Learning", Springer Science and Business Media LLC, 2023 Publication	1%
3	link.springer.com Internet Source	<1%
4	www.iieta.org Internet Source	<1%
5	"Proceedings of 4th International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications", Springer Science and Business Media LLC, 2024 Publication	<1%
6	d-scholarship.pitt.edu Internet Source	<1%
7	pure.ulster.ac.uk Internet Source	<1%

8	www.hindawi.com Internet Source	<1 %
9	Ritu Agarwal, Tanupriya Choudhury, Neelu J. Ahuja, Tanmay Sarkar. "Hybrid Deep Learning Algorithm-Based Food Recognition and Calorie Estimation", Journal of Food Processing and Preservation, 2023 Publication	<1 %
10	diva-portal.org Internet Source	<1 %
11	Niannian Wang, Xuefeng Zhao, Zheng Zou, Peng Zhao, Fei Qi. "Autonomous damage segmentation and measurement of glazed tiles in historic buildings via deep learning", Computer-Aided Civil and Infrastructure Engineering, 2019 Publication	<1 %
12	Ritu Agarwal, Tanupriya Choudhury, Neelu J. Ahuja, Tanmay Sarkar. "IndianFoodNet: Detecting Indian Food Items Using Deep Learning", International Journal of Computational Methods and Experimental Measurements, 2023 Publication	<1 %
13	Mahamudul Hasan, Nishat Vasker, Md Miskat Hossain, Md Ismail Bhuiyan, Joy Biswas, Mohammad Rifat Ahmmad Rashid.	<1 %

"Framework for fish freshness detection and rotten fish removal in Bangladesh using mask R–CNN method with robotic arm and fisheye analysis", Journal of Agriculture and Food Research, 2024
Publication

14	Submitted to De Montfort University Student Paper	<1 %
15	António Raimundo, João Pedro Pavia, Pedro Sebastião, Octavian Postolache. "YOLOX-Ray: An Efficient Attention-Based Single-Staged Object Detector Tailored for Industrial Inspections", Sensors, 2023 Publication	<1 %
16	"Proceedings of the 22nd Engineering Applications of Neural Networks Conference", Springer Science and Business Media LLC, 2021 Publication	<1 %
17	Kurnianingsih, Khalid Hamed S. Allehaibi, Lukito Edi Nugroho, Widyawan et al. "Segmentation and Classification of Cervical Cells Using Deep Learning", IEEE Access, 2019 Publication	<1 %
18	Submitted to KIET Group of Institutions, Ghaziabad Student Paper	<1 %

19	"Computer Vision", Springer Science and Business Media LLC, 2017 Publication	<1 %
20	Kong, F.. "DietCam: Automatic dietary assessment with mobile camera phones", Pervasive and Mobile Computing, 201202 Publication	<1 %
21	ia801805.us.archive.org Internet Source	<1 %
22	"Computer Vision – ECCV 2018 Workshops", Springer Science and Business Media LLC, 2019 Publication	<1 %
23	Haiqing Wang, Shuqi Shang, Dongwei Wang, Xiaoning He, Kai Feng, Hao Zhu. "Plant Disease Detection and Classification Method Based on the Optimized Lightweight YOLOv5 Model", Agriculture, 2022 Publication	<1 %
24	dokumen.pub Internet Source	<1 %
25	opsias.ias.ac.in Internet Source	<1 %
26	Almaghrabi, Rana, Gregorio Villalobos, Parisa Pouladzadeh, and Shervin Shirmohammadi. "A novel method for measuring nutrition	<1 %

intake based on food image", 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings, 2012.
Publication

27	Baby Shamini P, Nivetha R, Rituja Yadav R, Hemala R. "A Multi-Purpose Food Recognition System Using Convolutional Neural Network", 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), 2023 Publication	<1%
28	acikbilim.yok.gov.tr Internet Source	<1%
29	hdl.handle.net Internet Source	<1%
30	"Artificial Neural Networks and Machine Learning – ICANN 2016", Springer Nature, 2016 Publication	<1%
31	Submitted to Letterkenny Institute of Technology Student Paper	<1%
32	Shizhou Lu, Chenliang Ren, Jiexin Zhang, Qiang Zhai, Wei Liu. "A Novel Approach to Droplet's 3D Shape Recovery Based on Mask	<1%

R-CNN and Improved Lambert–Phong Model", Micromachines, 2018
Publication

---

33 Submitted to University of Queensland
Student Paper
<1 %

---

34 "Medical Image Computing and Computer Assisted Intervention – MICCAI 2019", Springer Science and Business Media LLC, 2019
Publication
<1 %

---

35 Mehdi Ghayoumi. "Generative Adversarial Networks in Practice", CRC Press, 2023
Publication
<1 %

---

36 José A. Romagnoli, Luis Briceño-Mena, Vidhyadhar Manee. "AI in Chemical Engineering - Unlocking the Power Within Data", CRC Press, 2024
Publication
<1 %

---

37 Xi Yang, Nannan Wang, Bin Song, Xinbo Gao. "Aurora Image Search With Saliency Deep Features", IEEE Access, 2019
Publication
<1 %

---

38 academic.oup.com
Internet Source
<1 %

---

39 yourpodiatrycentre.com.au
Internet Source
<1 %

40	Weixing Zhang, Chandi Witharana, Anna K. Liljedahl, Mikhail Kanevskiy. "Deep Convolutional Neural Networks for Automated Characterization of Arctic Ice-Wedge Polygons in Very High Spatial Resolution Aerial Imagery", Remote Sensing, 2018 Publication	<1 %
41	cdn.techscience.cn Internet Source	<1 %
42	www.mdpi.com Internet Source	<1 %
43	Katarzyna Kopczewska. "Applied Spatial Statistics and Econometrics - Data Analysis in R", Routledge, 2020 Publication	<1 %
44	Elham Nabizadeh, Anant Parghi. "Automated corrosion detection using deep learning and computer vision", Asian Journal of Civil Engineering, 2023 Publication	<1 %
45	encyclopedia.pub Internet Source	<1 %
46	vdoc.pub Internet Source	<1 %

47	"Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries", Springer Science and Business Media LLC, 2019 Publication	<1 %
48	Wenhao Lai, Feng Hu, Xixi Kong, Pengcheng Yan, Kai Bian, Xiangxiang Dai. "The study of coal gangue segmentation for location and shape predicts based on multispectral and improved Mask R-CNN", Powder Technology, 2022 Publication	<1 %
49	cris.maastrichtuniversity.nl Internet Source	<1 %
50	ruor.uottawa.ca Internet Source	<1 %
51	"International Conference on Innovative Computing and Communications", Springer Science and Business Media LLC, 2021 Publication	<1 %
52	Durgesh Kumar Mishra, Nilanjan Dey, Bharat Singh Deora, Amit Joshi. "ICT for Competitive Strategies", CRC Press, 2020 Publication	<1 %
53	Luciano da Fontoura Costa, Roberto Marcond Cesar Jr.. "Shape Analysis and Classification - Theory and Practice", CRC Press, 2019 Publication	<1 %

54	R. Dinesh Kumar, E. Golden Julie, Y. Harold Robinson, S. Vimal, Sanghyun Seo. "Recognition of food type and calorie estimation using neural network", The Journal of Supercomputing, 2021 Publication	<1 %
55	Simran Gupta, Md. Rahad Islam Bhuiyan, Sadia Sultana Chowa, Sidratul Montaha, Rashik Rahman, Sk. Tanzir Mehedi, Ziaur Rahman. "Enhancing Autism Spectrum Disorder Classification with Lightweight Quantized CNNs and Federated Learning on ABIDE-1 Dataset", Mathematics, 2024 Publication	<1 %
56	Ying-Chieh Liu, Djeane Debora Onthoni, Sulagna Mohapatra, Denisa Irianti, Prasan Kumar Sahoo. "Deep-Learning-Assisted Multi-Dish Food Recognition Application for Dietary Intake Reporting", Electronics, 2022 Publication	<1 %
57	Yu-Jin Zhang. "A Selection of Image Analysis Techniques - From Fundamental to Research Front", CRC Press, 2022 Publication	<1 %
58	core-cms.prod.aop.cambridge.org Internet Source	<1 %
59	mdpi-res.com Internet Source	

<1%

60	universe.roboflow.com Internet Source	<1%
61	worldwidescience.org Internet Source	<1%
62	www.uyik.org Internet Source	<1%
63	Fotios K. Konstantinidis, Vasiliki Balaska, Symeon Symeonidis, Dimitrios Tsilis et al. "Automating dairy production lines with the yoghurt cups recognition and detection process in the Industry 4.0 era", Procedia Computer Science, 2023 Publication	<1%
64	Jamalia Sultana, Benzir Md. Ahmed, Mohammad Mehedy Masud, A. K. Obidul Huq, Mohammed Eunus Ali, Mahmuda Naznin. "A Study on Food Value Estimation From Images: Taxonomies, Datasets, and Techniques", IEEE Access, 2023 Publication	<1%
65	Pedram Azad, Tamim Asfour, Ruediger Dillmann. "Combining Appearance-based and Model-based Methods for Real-Time Object Recognition and 6D Localization", 2006	<1%

IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006
Publication

66 Rakib Ul Haque, Razib Hayat Khan, A. S. M. Shihavuddin, M. M. Mahbubul Syeed, Mohammad Faisal Uddin. "Lightweight and Parameter-Optimized Real-Time Food Calorie Estimation from Images Using CNN-Based Approach", Applied Sciences, 2022 <1%
Publication

67 Usha Sakthivel, Nagarajan Susila. "Applied Learning Algorithms for Intelligent IoT", CRC Press, 2021 <1%
Publication

68 "Proceedings of Fifth International Conference on Computer and Communication Technologies", Springer Science and Business Media LLC, 2024 <1%
Publication

69 Hailiang Li, Yongqian Huang, Zhijun Zhang. "An Improved Faster R-CNN for Same Object Retrieval", IEEE Access, 2017 <1%
Publication

70 Parth Poply, Angel Arul Jothi J. "An Instance Segmentation approach to Food Calorie Estimation using Mask R-CNN", Proceedings of the 2020 3rd International Conference on <1%

Signal Processing and Machine Learning, 2020

Publication

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		